A Review of Tax Revenue Forecasting Models for the Scottish Housing Market

ECONOMY AND LABOUR MARKET

social research

A REVIEW OF TAX REVENUE FORECASTING MODELS FOR THE SCOTTISH HOUSING MARKET

Alma Economics

Scott Cameron
Nick Spyropoulos
Aliya Saied-Tessier
Sandu Gusanu
Napoleon Barkis

Contents

1. Context and scope

This report was commissioned by the Scottish Government in response to recommendations by the Scottish Fiscal Commission to explore model options for forecasting the housing market for Land and Buildings Transaction Tax (LBTT).[1] The research provides a comparative evidence base for forecasters to decide whether to change the current approach, and how to proceed should they decide to do so.[2] The review's specification has been jointly agreed by the Scottish Government and the Commission.

We focus on the residential tax base (that is, average prices and transactions volumes). Techniques for assessing the housing price distribution and applying tax rates to the base are addressed to a lesser extent. That said, the models and extensions we review can be applied to the wider economic and fiscal forecasting framework.

Budget forecasting is important for identifying trends that could affect the ability of the government to deliver its policy goals both in the immediate future and over the longer term. For this reason, it is important that forecasts be accurate. But the requirements of public sector forecasting models go beyond accuracy. They should also lend themselves to intuitive and transparent communication of revisions to stakeholders. Model developers must also be aware of practical constraints such as the availability of reliable and timely data, and should choose models that require appropriate resources to develop, run, and maintain. We attempt to capture these considerations in our evaluation. Our methods and criteria are described in Section 2.

Section 3 provides formal model assessments for eight general classes of models that have an evidence base of application to the housing market or may offer modelling potential in the future. These are: 1) technical assumptions, 2) univariate time series approaches, 3) multivariate econometric models, 4) vector autoregressive models, 5) error-correction models, 6) large-scale macroeconometric models, 7) dynamic stochastic general equilibrium models, and 8) microsimulation models.

Section 4 describes models and techniques that could be integrated alongside the assessed model classes in the budget forecasting framework to improve forecasts. These include Bayesian techniques, dynamic factor modelling, computable general equilibrium models, techniques to anticipate turning points (booms and busts), and forecasting the tax take directly.

Models for public budgeting come with trade-offs that are not only difficult to balance, but often at opposition to one another. For example, models that perform well at predicting the path of future revenues may not be designed to accurately estimate the causal relationships between variables and may underperform for fiscal impact costings and risk assessments. To deal with these competing objectives, many practitioners in budget

---

[1] See recommendation 3.32 of the Scottish Fiscal Commission's Report on Draft Budget 2016-17, available at http://www.fiscalcommission.scot/media/1094/draft-budget-2016-17.pdf.

[2] The Scottish Government's forecasting approach has evolved over time. For example, the modelling approach used for Draft Budget 2016-17 is described here: http://www.gov.scot/Publications/2015/12/7589, while the most recent modelling approach, used for Draft Budget 2017-18, is described here: http://www.gov.scot/Publications/2016/12/6669.

institutions maintain a suite of different housing models: one or more for predicting tax revenue and social expenditure, one or more for policy analysis (fiscal impact estimates and distributional analysis), and one or more for producing the macroeconomic forecast. The Scottish forecasting framework may similarly be best served by more than one approach. Section 5 describes how practitioners use these models together in forecasting frameworks.

The practical issues unique to public sector forecasting can heavily influence model selection and are not always given weight in academic forecasting literature. For this reason, we also conducted a series of practitioner interviews on the practical matters of public forecasting. Responses are provided in Section 6.

Section 7 summarises the results of the review and briefly discusses the next steps, should development of a new approach be taken forward.

2. Methodology and evaluation criteria

To identify candidate approaches, we searched peer-reviewed journals and the research of public and private sector forecasters. We also interviewed experts on housing market forecasting and budget preparation in several OECD governments, independent fiscal scrutiny bodies, and central banks. We narrowed our evaluation to the model classes in Section 3 based on these initial findings. We then selected the literature that was most relevant to Scotland to support a systematic assessment of these model classes.

The criteria against which we evaluated each model class are listed in Table 1. We first assessed whether the model was likely to be suitable for the Scottish housing market. We divided the application criteria into two components: 1) forecasting, and 2) policy analysis. The distinction between these two objectives is important for model selection and is discussed in detail in Box 1.

Assessing accuracy is difficult without developing the models themselves and comparing their out-of-sample forecasting properties to the current approach. As will be shown, research suggests that the accuracy of different model classes depends on the unique circumstances under which the model is asked to perform (that is, the forecast period, geographical region, and position within the business cycle). We were able, however, to provide indicative evidence from comparative studies. Because models tend to have different forecasting properties over the short and medium term, we attempted to find evidence on performance in the first eight quarters and the last three years of the intended five-year forecasting horizon. This breakdown could be useful for combining models.

We then assessed several elements of the model's ease of communication, included whether the model can tell a convincing story for both its current path and revisions since the last forecast round, and whether it is likely to be transparent to stakeholders.

We also ranked the model on whether its data requirements are likely to exist and be met in a timely matter, and the likely use of resources the model will require to develop, run, and maintain over time.

Each criterion was assigned a summary score of *good*, *fair*, or *poor*, based on the most likely scenario we can foresee without carrying through the model development itself. Results may vary in practice.

We did not attempt to assign weights of relative importance to the criteria, and their interaction is not straightforward. For example, a model that is specified to easily attribute forecast errors to economic determinants (that is, it scores highly on the communication criterion) is likely to require forecasts of exogenous economic variables. Forecasting exogenous variables may require hiring more analysts (reducing the model's resources score) and may introduce additional uncertainty (reducing the model's accuracy score). Scores should therefore only be taken as indicative and should not be considered in isolation.

Box 1: forecasting models versus policy models

A first question to ask is: what will the model be expected to do? If it will be required only to forecast baseline revenues, a wide variety of forecasting approaches are possible. However, forecasting models in a public budget setting are often asked to do other analysis, broadly defined as *policy analysis*. Policy analysis includes:

- **Fiscal impact costing**. The model may be asked to assess the fiscal cost (that is, the increase or decrease in government revenues) of changes in housing market policy or changes in LBTT policy, such as rates and thresholds.

- **Scenario analysis**. The model may be asked to assess the impact on housing demand of changes to other economic and fiscal assumptions (for example, stricter financial regulation or higher social assistance for lower income households).

- **Risk analysis**. Budget forecasters often publish a table of fiscal sensitivities or risks to the outlook—that is, how a change in GDP, inflation, or other inputs may affect revenues. This is useful, for example, if a minister is interested in how LBTT revenues would change if the outlook for economic growth were to fall below the budget's assumption, or if the central bank were to miss its inflation target.

Policy and risk analysis requires models that have been developed to capture the underlying process that generates the data. This largely rules out, for example, univariate forecasting models that make no attempt to describe how the housing market is influenced by other economic and fiscal variables of interest (and therefore cannot use alternative assumptions to compute costings, scenarios, and fiscal sensitivities).

Forecasting and policy models place different emphasis on the intertemporal dynamics of the data (the way that past values influence current and future values). Forecasting models place the importance of observed intertemporal properties and dynamics above all else. Policy models, on the other hand, place importance on economic and statistical theory to drive the equation specification. By emphasizing theory, policy models sometimes miss out on potentially useful forecasting information in the dynamics of the data.

Table 1: Assessment criteria

Criteria		Description
Application	forecasting	This score reports whether examples of that model class are common in the academic and practitioner forecasting literature for housing prices and transactions. Further, model classes that can produce *ex ante* forecasts (that is, unconditional forecasts without auxiliary forecasts of exogenous variables) will score well. Models that require auxiliary forecasts of exogenous variables will score in the mid-range (fair) and models that are not particularly suited for forecasting the housing market for tax purposes will score poorly.
	policy	If a model is a poor choice for forecasting, it may still be useful for policy. The purpose of policy models ranges from costing tax changes, assessing government interventions, dynamic scoring (estimating the feedback effects of fiscal policy on the economy), and scenario analysis (changing assumptions such as immigration levels or the exchange rate). Models that are structural (have been specified with an economic theory in mind) will score well.
Accuracy	short run (quarters one to eight)	Forecast accuracy can be measured using several different forecast assessment statistics. Although we did not develop the models and test their accuracy ourselves, we attempted to find comparative studies and rank general forecast conclusions relative to other model classes. A good score means models generally performed better than their peers (and poor, worse). A fair score suggests that evidence was mixed. Because models have different properties over different horizons, we split this into two assessments: one for the short run (the first eight quarters of the outlook) and one for the medium run (years three to five of the five-year outlook).
	medium run (years three to five)	
Communication	story telling	Forecasters are often asked to explain their model's outlook to policymakers and other stakeholders, and explain any revisions since the previous forecast round. Models with a theoretical basis for the specification and direct causal interpretation will score well on this criterion. Models that are mostly a black box using historical statistical properties will score poorly. Additionally, the forecast for a budget line item does not only need to tell an intuitive story in its own right, it must also be presented and framed within a narrative of the government's views on the economy. The model should therefore also be capable of being integrated into the budget framework in a consistent manner. A model that is sufficiently specified to include economic determinants from the macroeconomic outlook will score well on this criterion.
	transparency	When a model's assumptions are clear and transparent, it reassures stakeholders that forecasts are credible and assists external scrutiny. This criterion rates whether the model lends itself to transparent reporting of assumptions and whether it can be scrutinised readily by someone with a general economics background (but not necessarily an advanced specialist degree). A simple model that limits the role of judgment will score well on this criterion. More complex models involving considerable judgment, or requiring advanced degrees to interpret will score poorly.

Data compatibility	Most Scottish data on housing and other potential explanatory variables is available on a quarterly basis back to at least 2003 or the mid-1990s, depending on the level of aggregation. This criterion assesses compatibility of the model with 50 to 150 observations of quarterly data in a timely fashion.[3]
Resources	The resources, or cost, of the model, are related to its complexity and the specialised background it requires. This criterion rates the number and type of analysts required to run and maintain the model, the learning curve, and the ability for the work to be split and allocated within a team. A good score implies the model requires few team resources in proportion to the housing market's overall importance to the budget. A fair score implies the model could be implemented with resources proportionate to the role of the housing market in the budget. A poor score implies the model would require significant investments and greatly exceed the relative importance of the housing market in the overall budget.

3. Model assessment

Table 2 lists the broad model classes that we cover.[4] We focus the review on model classes and methods within model classes that are likely to be appropriate for the housing market in Scotland in a public budgeting context.

We start by describing the simple assumptions that many practitioners use in place of elaborate models. These include using a rule of thumb, a growth accounting framework (decomposing the series into its main drivers and projecting them forward mechanically), or a consensus forecast (an average of external non-government forecasters).

We then introduce univariate time series models that predict future values of housing prices and transactions using only the past statistical properties of the series itself. These include models that use the tendency of a series to return to its path following shocks (ARIMA modelling), models that estimate a series' tendency to behave differently over

---

[3] Scottish Government data relating to housing supply (starts and completions) is available on a quarterly frequency by local authority back to 1996, and at a national level to 1980 (and to 1920 for annual completions). Registers of Scotland quarterly statistics on prices and transactions go back to 2003, and annual data to 1993. The ONS House Price Index is available on a monthly basis back to 2004 for Scotland, using the new methodology covering both mortgage and cash sales. ONS house price data under the previous methodology, which is limited to sales with a mortgage, is available for Scotland back to the early 1990s on quarterly basis and to 1969 on an annual basis. The ONS Experimental Index of Private Housing Rental Prices has monthly data for Scotland back to 2011, while the Scottish Government Private Rent Statistics has annual data by broad rental market area and property size back to 2010. Quarterly data on mortgage affordability in Scotland is available back to the early 1990s from the ONS and to the 1970s from the Council of Mortgage Lenders, although there are breaks in methodology and coverage. Quarterly National Accounts Scotland, which include residential dwellings under gross fixed capital formation on an experimental basis, are available back to 1998.

[4] We excluded several models from the analysis. We did not evaluate state-space models, because they can generally be represented as the univariate and multivariate formulations we cover. We did not cover non-linear models in depth beyond GARCH and threshold autoregressive models (non-linear models are geared toward asymmetrical data and data with outliers, and are ill-suited for multiple step-ahead forecasts). We also did not review non-parametric approaches such as neural network forecasting, as many do not have explicit models (that is, they are black boxes) and often result in specifications similar to the approaches we cover, but would do poorly according to many of our criteria. In practice, these approaches have met with limited success when applied to economic data (Stock (2002) suggests that the biggest excluding factor is small data sets in economics, whereas these methods are data-intensive).

different periods (regime-switching models), and models that forecast a series' volatility (GARCH).

Next, we review multivariate econometric models grounded in theoretical relationships between the housing market and other economic drivers (such as employment growth, household incomes, and population). These models attempt to predict both the path of the series and why it will take that path.

We also look at models that combine both time series and multivariate approaches in a multiple equation simultaneously determined framework: the vector autoregression approach.

We follow this with error-correction models, which were created to address technical issues with non-stationary time series data, and large-scale macroeconometric models, which use a combination of techniques to forecast the housing market as a key component of a system of equations for the economy as a whole.

We then look at theory-based models that use optimising agents (micro-foundations) to arrive at a forecast: dynamic stochastic general equilibrium models.

Finally, we assess microsimulation models that use survey data and tax return samples (or sometimes the entire universe of transactions) to build up a complete description of taxpayers in the economy.

Table 2: Assessed model classes

Description	Class
Assumption or rule based	Rules of thumb Growth accounting models External consensus forecasts
Univariate time-series	ARIMA GARCH Regime-switching models
Multivariate approaches	Multivariate regression models Vector autoregressive models Error-correction models Large-scale macroeconometric models
Theory-based micro-founded	Dynamic stochastic general equilibrium models
Policy-focused models	Microsimulation models

3.1　Forecasting by technical assumption

Practitioners often use approaches that are mechanical in nature, requiring little to no judgment or estimation of model parameters. We refer to these methods as technical assumptions. There were three general forms of technical assumptions that appeared frequently in budget backgrounders and practitioner interviews: 1) rules of thumb, 2) growth accounting models, and 3) an average of external forecasts.

Rules of thumb are simple approaches that are grounded in theory, have shown value in practice, or are chosen subjectively because forecasters have judged that it is not worth applying significant modelling effort (for example, if the series is negligible as a percentage of the overall budget or GDP). They can be applied to both prices and transactions. They often resemble techniques in other model classes, with the difference that they generally do not contain estimated parameters or are estimated mechanically. Rules of thumb can take a range of sophistication:

- holding the variable constant in the future based on its last observation
- projecting it forward using its simple historical average (the mean)
- using a constant trend growth assumption (for example, its historical average growth rate)
- using an exponential smoothing model
- a simple one-to-one growth relationship with other economic variables, such as GDP

Holding a variable constant based on its last observation would be appropriate if, for example, examination of the series suggests that it follows no predictable pattern, with long periods trending up or down—that is, it is a random walk. If, however, it seems to fluctuate randomly around a stable value, then it may make sense to use its historical mean, or a moving average (the sample over which the mean is estimated goes back a limited number of periods and moves along as more observations are added).[5] These rules may be appropriate for housing transactions if new housing development is restricted and population, real incomes, and demographics are stable.

Similarly, projecting the series with a constant growth rate could be appropriate if historically prices have grown at roughly the same rate as general CPI price inflation, or for transactions if there are few restrictions on development and population is growing steadily.

Exponential smoothing models resemble a moving average, but values further in the past are given decreasingly smaller weights in its calculation. Exponential smoothing

---

[5] Formally, a random walk model for housing prices, P_t, would be: $P_t - P_{t-1} = e_t$. A useful forecast for a random walk is $P_{t+1} = P_t$, provided the random shock, e_t, is zero, on average.

A moving average model is defined as:

$$P_{t+1} = \frac{1}{k} \sum_{i=t-k+1}^{t} P_i$$

where k is the number of periods over which the average is taken. For $k = 1$, the moving average is equivalent to using the last observation as the forecast. For $k = t$, the moving average uses all the observations and is equivalent to the historical mean (Makridakis et al., 2008).

techniques have been developed that can handle trending and seasonal data, and are easily applied in a push-button manner by spreadsheet programs and statistical software.[6]

Rules of thumb could also be based on rough economic relationships, such as assuming the nominal housing tax base grows with the growth of nominal GDP. Growing the tax base with the growth rate of nominal GDP is the same as assuming it grows by population, inflation and real incomes (productivity). Equivalently, this assumes the average consumer will spend the same proportional amount of their income on housing services over time

Rules of thumb can be decided by forecasters using basic descriptive statistics and judgment, or they can be institutionalised to impart a degree of independence by being imposed by an arm's-length body such as the Auditor General for Scotland.

Alternatively, forecasters could use a **growth accounting model** that decomposes prices and transactions into their main cost drivers. For example, Moro and Nuño (2012) assume that average house prices in a period (P_t) are directly related to construction costs. They decompose costs into the cost of capital (the interest rate on borrowing, R_t) and the cost of labour (the wage rate, W_t) in the construction sector relative to the rest of the economy, using a growth equation like the following:[7]

$$P_t = P_{t-1} \cdot \frac{R_t}{R_{t-1}} \cdot \frac{W_t}{W_{t-1}} \cdot (1 + x_t)$$

where x_t is a residual growth factor capturing growth in excess of capital and labour costs. All variables are expressed as a ratio of prices in the construction sector to prices in the general economy.

This framework can be used to project housing prices using forecasts taken from a macroeconomic model for R_t and W_t. Practitioners typically hold future values of x_t constant at its historical average, unless there is a strong reason to suspect otherwise. Accounting methods incorporate economic drivers, but often assume away business cycle considerations (such as short-run deviations of supply and demand from equilibrium). For this reason, they are generally referred to as *projections* rather than forecasts (Belsky, Drew, and McCue (2007) discuss this distinction). Accounting projections are nonetheless common in forecasting frameworks reported by practitioners, especially when the series is volatile and difficult to predict, or for projections beyond five years (for example, as part of long-term debt sustainability projections that assume markets are in equilibrium).

Finally, the Scottish forecasting framework could use an average of **non-government forecasts**. This would involve regularly surveying private sector banks, real estate industry firms and experts, think tanks, and universities for their outlook for the housing market. The individual forecasts would then be combined using a simple average for each year of the forecast. If not used to generate the housing market forecast itself, the survey could be used to generate exogenous economic assumptions for more sophisticated models.

---

[6] A common generalised exponential smoothing approach is the Holt-Winters method. See Makridakis et al. (2008) for more on its theoretical derivation and application.

[7] This has been simplified. The Moro and Nuño model allows for capital and labour to be weighted by their intensity in production (α_i), where i = {c,g} and c is the construction sector and g is the general economy. The full specification, in continuous time notation, is:

$$\frac{\dot{(P_{c,t}/P_{g,t})}}{(P_{c,t}/P_{g,t})} = \alpha_c \frac{\dot{R_{c,t}}}{R_{c,t}} - \alpha_g \frac{\dot{R_{g,t}}}{R_{g,t}} + (1-\alpha_c) \frac{\dot{W_{c,t}}}{W_{c,t}} - (1-\alpha_g) \frac{\dot{W_{g,t}}}{W_{g,t}} + \frac{\dot{(A_{g,t}/A_{g,t})}}{(A_{g,t}/A_{g,t})}$$

Assessment

Application (forecasting): good. Examples of rules of thumb come mostly from practitioner literature and interviews. Prior to the creation of the Office for Budget Responsibility (OBR), the UK budget was prepared in part by using a set of basic assumptions audited by the National Audit Office, including assumptions for trend GDP growth, unemployment, equity prices, oil prices, and tobacco consumption trends.[8]

The OBR uses a rule of thumb for its forecasts of the devolved Scottish LBTT and Welsh SDLT taxes, assuming they remain at a constant share of their forecast for the UK as a whole (OBR, 2016).

Practitioners reported that assuming a constant share of GDP is the "go-to" assumption for small taxes, or taxes for which data on fundamentals is limited.

The Scottish Government forecast housing prices in Draft Budget 2016-17 using a rule of thumb for the outer years of the outlook, interpolating between model results after the second year of the outlook to the historical average growth rate for prices in the fifth year. For transactions, the Scottish Government used a linear interpolation rule of thumb between the last historical value of the turnover rate to the historical average turnover rate imposed on the fifth year of the forecast horizon.

A growth accounting framework is used to project medium-run demand for homes based on fundamentals by the Joint Center for Housing Studies of Harvard University and presented most recently in Belsky et al. (2007). This approach models US housing demand with a simple accounting relationship based on three factors: 1) net household growth (itself projected using headship rates and immigration), 2) the net change in vacant units (calculated with demand for for-sale vacancies, for-rent vacancies, and second and occasional use homes—in turn projected by the age distribution of population, household wealth, and preferences), and 3) replacement of units lost on net from existing stock as a result of disaster, deterioration, demolition, and conversion to non-residential units. Although the framework is for new home construction, it could be extended to include turnover for existing homes for a projection of total transactions.

Belsky et al. also present an alternative and simpler approach to projecting total demand for new housing using the historical ratio of household growth to completions (the two most reliable housing data sources according to the authors).

If short-run dynamics are desired, McCue (2009) provides an extension of the Belsky et al. framework to compare the demand projections to actual supply to arrive at a short-run forecast of excess new supply and inventories. This excess (or deficit) supply measure could be used to introduce short-run cyclicality in prices using a multivariate regression model (discussed further in Subsection 3.3) while still being anchored at the end of the medium run by the accounting framework.

Growth accounting models may, however, be less suited to forecasting the UK housing market than the US market. There is convincing research that housing supply in the UK is inelastic (see Subsection 3.3), which suggests cyclical demand considerations could have a significant impact on prices.

---

[8] See Box C1 in Budget 2007, Economic and Fiscal Strategy Report and Financial Statement and Budget Report available at http://webarchive.nationalarchives.gov.uk/.

Examples of government forecasters using the private sector average include the OBR and HM Treasury (HMT), who, until December 2013, used an average of private sector forecasts for their 2-year ahead outlook for house price inflation (Auterson, 2014). Eighteen of the 38 external forecasters in HMT's October 2016 survey provided forecasts for housing price inflation for the UK as a whole for at least five quarters.[9]

Canada's federal Department of Finance uses the average of private sector forecasts for all key macro variables including real GDP growth, GDP inflation, real interest rates, and exchange rates. However, they use an internal macroeconometric model constrained to the consensus growth rates to decompose GDP into its components and income shares for fiscal forecasting, including housing prices and transactions. They have maintained the internal capacity for complete macro forecasting, but impose the private sector average as a basis for fiscal planning to "introduce an element of independence into the Government's fiscal forecast."[10]

Application (policy): fair. Neither of the three broad technical assumptions lend themselves to policy costings. However, there may be some scope to adjust assumptions to produce sensitivity tables and assess alternative assumptions—for example, if the growth rate of GDP or the consensus forecast of house price inflation were used, the rates could be adjusted by one percentage point and the consequences to the fiscal outlook could be reported. Alternative capacity for policy costings would need to be developed.

Accuracy (short run): fair. Forecasting by a technical assumption does not necessary mean sacrificing forecasting performance. On the strength of academic research on the unpredictability of many economic time series, practitioners are increasingly foregoing sophisticated forecasting techniques in favour of simple assumptions. For example, recent research such as Alquist and Vigfusson (2013) found that the common approach to forecasting the price of oil using the oil market futures curve cannot beat a simple no-change assumption. Their work has led many practitioners to abandon sophisticated oil price models.[11]

However, for the housing market, there is sufficient research to reject the idea that sophisticated models cannot improve upon technical assumptions. Researchers as early as Case and Shiller (1988) convincingly demonstrated that housing markets are not efficient (that is, they do not follow a random walk and observant investors can earn returns above a safe rate). This suggests housing time series are predictable. For example, if they experience one year of above-average growth, the next year tends also to be above average. Therefore, simple rules are unlikely to do well in cross-model comparisons based on accuracy.

Moro and Nuño (2012) tested their growth accounting framework empirically over the 1980s to late 2007 in four countries: the UK, US, Spain, and Germany. They found that it provides a useful description of housing price movements only in the US.

---

[9] See Forecasts for the UK economy: various months, available at: https://www.gov.uk/government/collections/data-forecasts.

[10] See Backgrounder—Canadian Economic Outlook, available at: http://www.fin.gc.ca/n16/data/16-025_1-eng.asp.

[11] For example, see the discussion on oil prices in the Bank of Canada's July 2014 Monetary Policy Report, available at: http://www.bankofcanada.ca/wp-content/uploads/2014/04/mpr-2014-04-16.pdf.

There is considerable evidence that averaging forecasts, as in the consensus approach, can provide superior forecasting performance (for example, Meulen et al. (2011) and Granziera, Luu, and St-Amant (2013)). However, in the case of many economic and fiscal variables, governments may have more timely and accurate information than private sector forecasters (for example, real-time VAT receipts). Private sector economic forecasters may also have biases unique to their circumstances. For example, there may be a bias in the public forecasts of private sector investment banks as a result of financial incentives (few would be enthusiastic to invest if the economic outlook is grim) and for mechanical reasons related to lags in recognising downturns (Batchelor, 2007).

Accuracy (medium run): fair**.** The medium run should benefit from forecast averaging or simple anchors based on high-level economic trends or the variable's history. However, there is some evidence that suggests otherwise. Tsolacos (2006) evaluates a consensus forecast of real estate rents (a returns index) and finds that while the consensus forecast for rents is best at a one-year forward horizon, simple time series approaches and regression models with interest rates outperform the consensus forecast two years out.

Communication (story telling): fair. Technical assumptions vary in their ability to tell a story—for example, using the historical mean or average growth rate would not reflect economic fundamentals, but growing prices or transactions with GDP may capture general economic trends. Growth accounting models allow broad trends to be discussed, but may not be able to explain short-run changes related to the business cycle.

The private sector average can tell a story fitting both overall economic trends and trends in the housing market, depending on how detailed the survey is; however, budget-to-budget revisions may be impossible to explain.

Communication (transparency): good. Presenting and explaining technical assumptions is relatively straightforward. All three approaches can be made independent and transparent. If the consensus forecast is viewed as a form of externally-imposed rule, it is transparent. However, the underlying methodology and assumptions that non-government forecasters use to produce the forecast would typically not be available.

Data compatibility: fair. Generally, technical assumptions have few data requirements, and the ones that do (such as for growth accounting models) are at a high level that will be available to Scottish forecasters. However, there are some limitations that reduce this score to *fair*.

First, using a private sector average may prove challenging for a forecast of the Scottish market. Relatively few institutions produce Scottish forecasts, and even fewer offer detailed forecasts down to the residential housing sector level, especially for a five-year forecast horizon. However, there are surveys of professional housing forecasters that may satisfy the requirements.[12]

The OBR suggests that although the consensus forecast was effective for communicating with stakeholders, it was abandoned largely because the data was not timely and definitions were problematic. This is summarised by Auterson (2014):

> [the consensus forecast] had the advantage of being simple and transparent, but the disadvantage that there can be a significant lag between new information becoming

---

[12] For example, the Investment Property Forum (IPF) has a number of survey products available: http://www.ipf.org.uk/

available and external forecasts being updated, collated and published. This problem is particularly apparent when house price inflation is changing rapidly. Another drawback was that external forecasts reference a number of different house price indexes, meaning only a subset are directly relevant to the ONS house price series we forecast. (p. 1)

Finally, rules of thumb that rely on basic statistics such as historical means do not work well with trending data, seasonality, or level shifts (Makridakis, Wheelwright, and Hyndman, 2008). This will largely rule out these approaches for Scottish data in levels, but may be suitable for transformed data. This will require further evaluation.

Resources: good**.** Technical assumptions are cost effective and require few analytical resources and little to no specialist skills to apply. However, internal modelling capacity may still be required for policy analysis. Technical assumptions are easily estimated or imposed in spreadsheets and statistics software packages.

Our assessment of forecasting by technical assumption is summarised in Appendix Table A1.

3.2 Univariate time series approaches

Univariate time series models predict housing prices and transactions based solely on their own statistical properties, particularly the relationships of the variable with its values in the past (that is, the correlations and partial correlations estimated by regressing the variable on time-lagged values of itself). For example, if residential prices rose more quickly this year than their trend, it may be likely that they will also rise more quickly next year. These properties can be used to predict future values of prices and transactions without considering the wider economy.[13]

We look at three univariate forecasting approaches. First, the tendency for price or transaction shocks to dampen or persist can be modelled using a generalised approach called ARIMA modelling. Second, the behaviour of a series may change over different periods, such as when it is trending up or down (for example, during a housing boom or bust). This type of behaviour can be modelled using regime-switching methods. Third, univariate models can predict another property of the series: its volatility. This can be modelled using a GARCH process and may be useful in forecasting the risk to the budget outlook.

ARIMA

The general form of a univariate time series model is the autoregressive integrated moving average (ARIMA) model, attributed to Box and Jenkins (1976), who first described the technique in detail and created a systematic approach for its estimation.[14]

---

[13] The rules of thumb based on moving averages and exponential smoothing models described in Subsection 3.1 are special cases of the model class described here. However, in the context of this section they are not necessarily fixed and mechanical, and require greater analyst effort and judgment in their specification.

[14] The AR component of the model was first introduced by Yule (1926). Slutzky (1937) described MA models. Wold (1938) combined the AR and MA models and has shown that ARMA processes can be used to model all stationary time series as long as the appropriate number of AR terms and the number of MA

ARIMA modelling attempts to capture two basic types of time series behaviour and their combination:

1. Autoregression
2. Moving average

The autoregressive (AR) component presumes that housing prices or transactions are a function of lagged values of themselves. That is, future values can be forecast with current and past values by estimating the correlation of the series' value in time *t* with its values in time *t-1*, *t-2*, *t-3*, etc. The autoregressive model has the following general form (as given by Enders (2014)):

$$x_t = c + a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_p x_{t-p} + e_t$$

where *c* is a constant (often not required), e_t is a random shock with constant variance, and *p* is the last lagged value that affects the current realisation.

The moving average (MA) component presumes that housing prices or transactions are a function of random surprises in previous years—that is, the difference, or errors, between the model and actual observations as time advances.

The general form of an MA process is:

$$x_t = m + b_1 e_{t-1} + \cdots + b_q e_{t-q} + e_t$$

The MA in an ARIMA process is a different concept than the moving average smoothing techniques discussed in Subsection 3.1. Here it is a behaviour of the error term of the model—a similar averaging process, but applied to the forecast errors of the series, instead of its past values.

AR and MA models can be combined in a general univariate time series model—the ARMA model. A simple ARMA model that depends only on its value and error in the previous round takes the form:

$$x_t = c + \varphi_1 x_{t-1} + e_t - \theta_1 e_{t-1}$$

An ARMA model can be used only if the series is *stationary,* a condition unlikely to be met by housing market variables (an overview of Scottish housing market data along with a definition of stationarity is given in Box 2). For example, prices are likely to increase each year to some extent along with other prices in the economy (general price inflation). In this case, the mean (average) of house prices would grow over time. An ARMA process on the level of house prices would generally under-predict prices. But the series can be transformed to be stationary. If the trend is steady and predictable (in the inflation example, prices may follow the Bank of England's inflation target), the variable can be made stationary by detrending the data (regressing the series on a time trend). If, however, the data has a stochastic trend (that is, it is unpredictable), the variable must be

---

terms are properly specified. Box and Jenkins (1976) popularized the ARIMA models and created a structured model selection process.

transformed by differencing.[15] The latter case means the variable is integrated, which is the abbreviated letter *I* in ARIMA.[16]

Regime-switching models

Regime-switching models allow parameters to take different values in each of a fixed number of historical intervals. A change in the model's parameters may arise from a range of causes, such as changes to monetary policy or changes to the exchange rate regime (Stock, 2002).

Regime-switching models fall into two categories: *threshold* models and *Markov switching* models. In threshold models, regime changes arise from the behaviour of an observed variable relative to some threshold value. These models were first introduced by Tong (1983). They are formulated in general terms in Stock (2002) as:

$$y_{t+1} = d_{\mathrm{t}}\alpha(L)y_t + (1 - d_t)\beta(L)y_t + e_{t+1}$$

Where *α(L)* and *β(L)* create coefficients and lags of the variable against which they are multiplied, and d_t is a non-linear function of past data that switches between the parameter regimes *α(L)* and *β(L).* Different functional forms of d_{t} determine how the model transitions between regimes.

In Markov-switching models, the regime change arises from the outcome of an exogenous, unobserved, discrete random variable assumed to follow a Markov process (that is, the history of the variable does not offer any more information about its future than its current value).[17] The general form of a Markov-switching model is similar to the threshold model but the function d_t represents the unobserved Markov process variable.

Forecasting volatility

The variance of an economic series is a measure of risk. If forecasters are interested in forecasting the risk to the LBTT outlook, or quantifying the variance of housing prices or transactions at different points in time, the series' own history can be used to forecast the variance. This is known as generalised autoregressive conditional heteroskedastic (GARCH) modelling, where heteroskedastic means that the variable's volatility is not constant over time.

GARCH models were developed first by Engle (1982) and later refined by Bollerslev (1986). GARCH may be thought of as an extension of the ARIMA method that forecasts using the typical time series behaviour of both the values of a series, and its variance.

The following equation is a simplified GARCH model, based on Enders (2014):

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_p u_{t-p}^2 + \sigma_t^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2$$

where u is the error term of a simple *AR(p)* process and σ_t^2 is the conditional variance of u which depends on the information available at time *t-1*.

---

[15] The first difference of x_t (Δx_t) is formed by subtracting the function in the previous period from the function in the current period: $\Delta x_t = x_t - x_{t-1}$

[16] A variable is integrated of order I(1) if differencing renders it stationary. It is integrated of order I(2) if it must be differenced twice to render it stationary.

[17] Formally, a Markov process's future and past are independent, conditional on its current state.

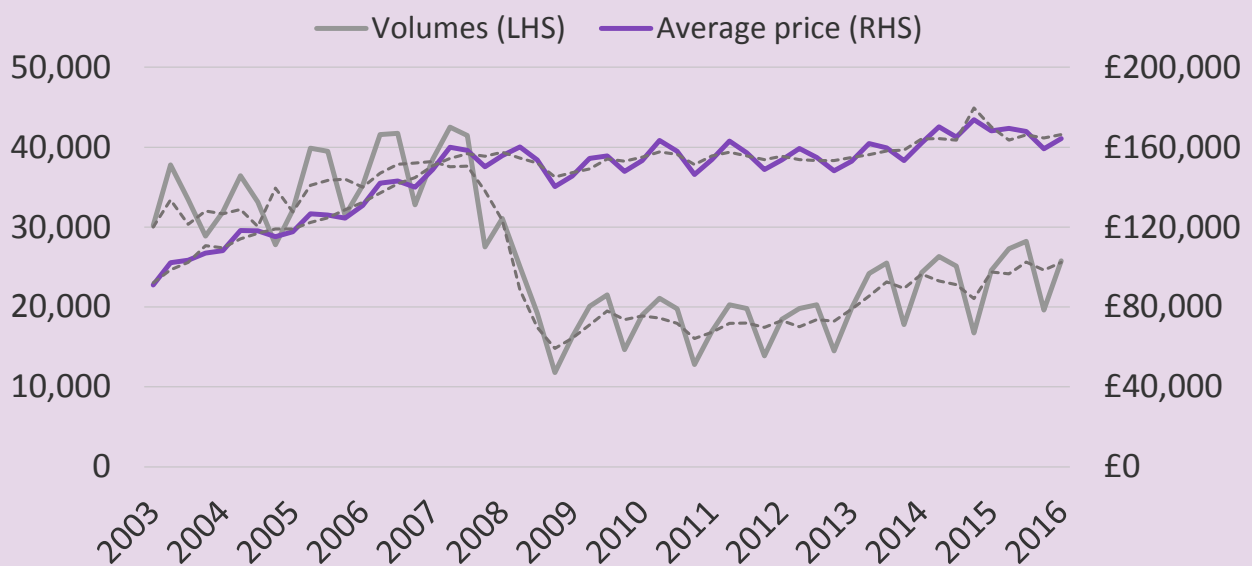**Box 2: Overview of Scottish residential prices and transactions**

Figure B1: Scottish residential housing prices and transactions

Figure B1 plots Scottish housing prices and transactions volumes from the Registers of Scotland, along with the seasonally adjusted series using the X-13ARIMA-SEATS procedure of the US Census Bureau.

Housing prices and transactions contain a trend and seasonal pattern. Both series were affected by (and in turn affected) the downturn following the global financial crisis. Between late 2007 and early 2009 housing transactions collapsed to around a third of their pre-crisis levels. The fall in transactions coincides with a structural break in trend prices, ending the strong growth preceding the crisis.

A main concern when specifying a forecasting model is whether a series is stationary—that is, it has a constant mean and variance over time. Both smoothed series show a clear trend over time and transactions seasonality seems to widen after 2013. This suggests the data is non-stationary (and indeed housing market data in the UK and abroad is routinely found to be non-stationary (see Barari et al. (2014), among others). To use many of the forecasting techniques evaluated in our review, the data would need to be transformed, either by deseasonalising if they are found to be stationary in level terms (for example, through using dummy variables to capture seasonality), detrending if they are found to be stationary around a constant time trend, or differencing if they are found to be stationary around a stochastic trend (this can include seasonal differencing, that is, using the annual growth rate for each quarter). Alternatively, special techniques can be used to maintain the model in levels (see Subsection 3.5—error-correction models).

Modelling is further complicated by policy changes such as the stamp duty holiday between September 2008 and December 2009, the introduction of graduated "slice" tax rates (similar to how personal income taxes are structured) to replace the previous "slab" rates on 4 December 2014, and the transition from the reserved stamp duty land tax (SDLT) to the devolved land and buildings transaction tax (LBTT) on 1 April 2015.

Assessment

Application (forecasting): good**.** ARIMA, regime-switching, and GARCH models have been applied to forecasting housing prices and transactions in a wide range of academic literature.

Potentially useful applications of ARIMA models to prices include Barari, Sarkar, and Kundu (2014), Stevenson and Young (2007), and Crawford and Fratantoni (2003), among others. For applications to volumes measures for transactions and housing supply, see Fullerton, Laaksonen, and West (2001).

Among practitioners, ARIMA modelling was the approach used by the Scottish Government in Draft Budget 2016-17 for forecasting average house prices in the first two year of the outlook, and for the whole five-year forecast horizon in Draft Budget 2017-18. Outside of Scotland, it is not widespread, but is often used as a benchmark comparison. ARIMAs are, however, used widely for economic forecasting and for revenues that are a small percentage of the overall tax take, or that do not have a tax base that lends itself to modelling directly.

Although we heard of no regime-switching models applied among practitioners, they are popular in academic research. For example, Park and Hong (2012*)* observed that monthly indexes of the US housing market are released at the end of the subsequent month. This, in turn, creates a month-long standstill in making judgements about the housing market. They exploit this interval to show how Markov-switching models can be used to promptly forecast the prospects of the US housing market within the standstill period.

Enders (2014) suggests that GARCH models are particularly suited to asset markets such as the housing market, as prices should be negatively related to their volatility (if market participants are risk-averse). Further, referring again to Box 2, it seems that the variance of house transactions in Scotland changes over time. The housing forecast may therefore benefit from GARCH modelling

Univariate models can produce *ex ante* forecasts without needing to be conditioned on auxiliary forecasts of exogenous variables (that is, they can forecast using only historical information available at the time of the forecast).

Application (policy): poor**.** All three methods of univariate forecasting are ill-suited for scenario analysis and fiscal impact costing, as they are not specified with explanatory variables to assess the impact of different economic assumptions and risk scenarios on the housing market.[18] ARIMA and GARCH models may have some limited use in risk assessments: ARIMA models can assess how exogenous shocks in one period are transmitted to future house prices and transactions in the future, and GARCH models may be able to improve upon estimates of annual revenue at risk.

Although of limited use for policy analysis themselves, univariate models may be a useful component of a broader policy analysis approach, such as for projecting components of the price distribution for the application of tax rates.

---

[18] In this context, policy analysis means estimating the impact of changes in housing market policy on the wider economy and vice versa. The direct impact of policy changes relating to tax rates and thresholds on tax revenue can still be assessed by combining univariate models which forecast average prices and transactions with house price distribution models, such as using a standard distribution (e.g. the log-normal distribution used in Scottish Government forecasts) or micro-simulation techniques/micro-databases, as discussed in Subsection 3.8

Accuracy (short run): good. ARIMA models showed mixed but broadly positive performance in out-of-sample forecasts, and in many cases, outperformed the more sophisticated models they were compared against. However, researchers are generally quick to assert that performance is dictated by the specific regions and time periods under study.

For example, the ability of ARIMA models to forecast Irish housing prices was evaluated by Stevenson and Young (2007) for the period 1998 to 2001. They found that ARIMA models provided more accurate forecasts compared to two other models—a multivariate regression and VAR model—on five forecasting accuracy measures: mean error, mean absolute error, mean squared percentage error, and error variance.

Lis (2013) estimated ARIMA models and other classes for rolling forecasts of the Canadian real estate markets in Vancouver and Toronto. Lis found that no single forecasting model performed best in all situations, but rather concludes that a forecasting approach should be chosen using detailed diagnostics for each series and time under study.

Brown et al. (1997) compared the ability of a regime-switching to predict house prices in the UK in the 1980s and 1990s to an error-correction model, an AR and a VAR model. They found that the regime-switching model performed the best in out-of-sample forecasts.

Meulen et al. (2011) constructed a unique price index using online real estate listings to control for different housing characteristics. They estimated a simple autoregressive model, as well as a VAR that incorporated information about macroeconomic trends. They found that the macroeconomic variables only slightly reduced the forecast errors compared to the naïve autoregression.

Maitland-Smith and Brooks (1999) found that Markov switching models are adept at capturing the non-stationary characteristics of value indexes of commercial real estate in the US and UK. The researchers found that it provided a better description of the data than models that allowed for only one regime (for example, a simple AR model).

Barari et al. (2014) estimated an ARIMA model and two regime-switching models on a 10-city composite S&P/Case-Shiller aggregate price index they created for seasonally-adjusted monthly data from January 1995 to December 2010. They found that the ARIMA model performs as well as the regime-switching models in out-of-sample forecasts.

Accuracy (medium run): fair. Forecast tests for univariate models in the academic literature are rarely performed more than two years into the future. The length of the useful forecast horizon is determined by the speed of decay, which is rarely significant beyond 8 quarters; however, when specified to decay to a simple trend, they may prove sufficient for the medium run.

Larson (2011) provides useful benchmark comparisons between univariate and multivariate models for three years into the future, finding that for Californian housing prices univariate time series models were outperformed by multivariate over the 1970s to late 2000s.

Communication (story telling): poor. Univariate time series models do not generally offer a direct causal interpretation of coefficients and can be difficult to communicate. That is, they predict what will happen, not why (Hyndman and Athanasopoulos, 2012). However, a univariate equation need not be entirely atheoretical, as a complex system of multivariate

explanatory equations can often be transformed into a univariate reduced-form equation (Enders, 2014). In that manner, a univariate time series estimated by Ordinary Least Squares (OLS) can capture the theoretical relationships of a wide assortment of underlying economic relationships. Nonetheless, the signs and magnitudes lose much of their ability to be interpreted, and the structural properties are impossible to recover from the final estimated equation.

Communication (transparency): fair. Equations and estimated coefficients would need to be published frequently, as specifications and estimates are likely to change with each addition of new or revised data. Fiscal sensitivity tables could not be estimated and published to provide a check on model revisions given economic developments. However, their relative simplicity lends them some merit, as scrutinizers with a general economics background would largely be able to understand and test the assumptions.

Data compatibility: good. The key advantage of this method is that it does not place a large burden on data collection—only historical data for the variable being modelled is required. Newton (1988) recommends a minimum of 50 observations for ARIMA modelling, which is in-line with the given history of reliable and detailed Scottish data. Univariate time series models are therefore well-suited to the available data for the Scottish housing market.

Resources: good. ARIMA models are an accessible forecasting model for small teams with limited technical background. Most software packages and forecasting guides have detailed procedures for the Box-Jenkins methodology that can guide the model selection procedure.

Our assessment of univariate time series models is summarised in Appendix Table A2.

3.3 Multivariate regression models

Rather than rely only on the past statistical behaviour of housing prices and transactions, forecasters can look for other factors that influence the housing market, such as interest rates and population. Models that include other explanatory variables are known as multivariate regression models.[19]

These models often use simple regression techniques such as OLS to predict future values of prices and transactions. They are similar to cross-sectional econometric analysis, except that explanatory variables are a function of time, and the estimated parameters can vary over time. For example, a simple multivariate forecast of housing prices may take the form:[20]

$$P_t = \beta_0 + \beta_1 \Delta GDP_t + INT_t + e_t$$

where: ΔGDP_t = the change in gross domestic product, representing general sentiment about the strength of property markets and the wider economy.

---

[19] Multivariate econometric regression models are often called behavioural models. When looking at average house prices and total transactions, behaviour refers to the relationships at an aggregate, economy-wide level. This contrasts with behavioural equations at the individual household level, for example modelling the impact of mortgage interest rates on decisions on house size. Modelling of the aggregate household sector often assumes that these same decisions can be observed in macroeconomic relationships.

[20] This is a heavily stylized model based on equations (2) and (5) of Tsolacos (2006).

INT_t = some measure of interest rates, capturing the cost of borrowing, the discount rate on future housing benefits, and the risk-free rates on capital gains for competing investments, among others.

The variables to include and the model specification are guided by economic theory, particularly the interaction of demand for housing by households and the supply of housing by land and building developers.

Importantly, forecasting with multivariate models requires forecasts of the future values of explanatory variables that will need to be provided by exogenous forecast models (Subsection 3.4 considers a technique where the future values of explanatory variables can be endogenously forecast).

Economic theories of the housing market

Multivariate regression equations in academic literature and practitioner research was most often based on **asset-pricing theory**. This approach was presented in an influential 1984 paper by Poterba. Asset-pricing models base the level of housing prices on an equilibrium concept with the 'income' that houses generate.[21] This income includes the value of housing services to the owner-occupier.

In Poterba's words:

> A rational home buyer should equate the price of a house with the present discounted value of its future service stream. The value of future services, however, will depend upon the evolution of the housing stock, since the marginal value of a unit of housing services declines as the housing stock expands. The immediate change in house prices depends upon the entire expected future path of construction activity. The assumption that the buyers and sellers of houses possess perfect foresight ties the economy to this stable transition path and makes it possible to calculate the short-run change in house prices which results from a user cost shock. (1984, p. 730)

From another perspective, these models assume an arbitrage opportunity between buying and renting: if the cost of renting a house is lower than the expected cost of buying and occupying an equivalent house (the *user cost* of housing), then owners will sell and rent instead, increasing the supply of homes for sale and reducing vacant rentals until rents and user costs converge. A simplified specification of the user cost of housing could look like the following, based on Auterson (2014) and others:

$$C_t = P_t * (i_t + \tau_t + \delta_t + m_t - g_t)$$

where P_t = the real price
i_t = the net mortgage rate
τ_t = the property tax rate
δ_t = depreciation
m_t = maintenance and repair expense

---

[21] The asset approach can be quite involved to implement in practical terms. There are simpler alternatives. For example, the **housing affordability approach** models house prices by examining the current affordability of housing relative to its long-run trend. The measure could be the ratio of per capita household income to the average house price or the ratio of mortgage interest payments (and capital repayments) to disposable income. If these ratios get too out of line with their historical trend, there should be pressure to move back toward the trend (Brooks and Tsolacos, 2010).

g_t = expected capital gain (inflation plus the real expected price increase)

The market clearing rental rate is taken as a function of housing supply, or other variables such as real incomes and demographics.

$$R_t = f(H_s, Y_t, ...)$$

By setting the rental rate equal to the user cost equation and manipulating the resulting equation, growth in housing prices can be specified as an inverted demand function: [22]

$$Real\ Price\ change_t = -R_t + P_t * (i_t + \tau_t + \delta_t + m_t - g_t^*)$$

where g^* is now the real expected capital gain. Auterson provides a description of how this model may be implemented in practice in the UK, including detailed data sources.

Implementing multivariate models in a forecast can involve simple applications of linear regression analysis with single equation specifications, for example an equation with price on the left-hand side (dependent variable). Models could also describe the housing market more generally with several equations, having price and transactions simultaneously determined, each driving one another within a feedback loop. The latter could include both structural and reduced-form systems of equations.[23]

Multivariate models could also be constructed using a bottom-up or top-down approach. In the bottom-up approach, relationships are estimated by looking at the factors that influence a house's value (such as the number of bedrooms, detached versus higher-density units) or real estate markets in different cities, and then aggregating to the national level based on population weights and construction trends. Models could also be applied separately for land and structures or for new construction versus existing turnover. Typically, however, structural relationships are examined at the top-down level. In a top-down approach, the relationship is examined at the national (aggregate) level using a model that relates aggregate average house price growth to macroeconomic variables such as growth in real incomes and employment.

Assessment

Application (forecasting): fair. There are many examples of multivariate regression models applied to forecasting housing prices and transactions in the UK and abroad. For example, Dicks (1990) extended the multiple equation demand and supply models of earlier researchers in the US to the UK market to forecast house prices for new and second-hand dwellings, as well as housing completions and the uncompleted stock of dwellings.

There is also a strong base of multivariate model research for the US market, owing to the richness and ease of access to data. In an influential paper, Case and Shiller (1990) pooled data across four US cities using OLS regression to examine how prices evolved based on explanatory variables such as the change in per capita real income, employment

---

[22] Usually in economics, markets are modeled as the quantity demanded at different prices. An inverted demand function solves the system differently to show the price level for a specific quantity demanded.

[23] Structural models allow contemporary effects of variables on one another, whereas reduced-form equations express dependent variables as a function only of lagged values of itself and other independent variables. A reduced-form model is usually derived from a structural set of equations through algebraic manipulation and makes estimation easier.

expansion, the change in the adult population, and changes in housing construction costs.[24] They estimate several specifications of forecasting models that prove to have significant forecast accuracy.

This approach is also common among practitioners. For example, the OBR's model as presented in Auterson (2014) describes the multivariate model they use to forecast the housing base for Stamp Duty Land Tax and for constraining the housing sector of the macroeconomic model. They model rental prices using real incomes, housing starts, and demographics, modifying the equation above to include an estimate of credit conditions, *mrat*, as follows:[25]

$$\ln(P_t) = \ln(R_t(RY_t, Hs_t, demographics) - \ln(i_t + \delta - g + f(mrat))$$

The OBR's model is based on several papers by Meen (2013, 2009, and 1990, among others) who has undertaken a wide range of research on the UK housing market.

Multivariate models of the housing economy can be used as auxiliary models outside of the main macroeconomic forecasting model, and their outputs can be imposed on the macroeconomy model, or, where accounting concepts are different or aggregated accounting identities need to hold, are used to apply judgement to the central model's equations (for example, see Kapetanios, Labhard, and Price, 2008).

Because multivariate models are conditioned on exogenous explanatory variables, they cannot produce *ex ante* forecasts. This reduces their score compared to models that can produce *ex ante* forecasts.

Application (policy): good. Multivariate econometric models are particularly relevant for policy assessments and fiscal impact costings. They can include a wide range of variables representing government policy and other explanatory variables that can be changed to estimate the cost of policy or to evaluate alternative assumptions. Makriditis et al. (2008) suggests governments have "few alternatives other than econometric models if they wish to know the results of changes in tax policies on the economy (p. 301)."

Accuracy (short run): fair. Single equation models can be tailored to fit the historical data perfectly, if enough explanatory variables are added. This is not necessarily an indication of useful forecasting performance, however, and in fact can lead to the opposite—overfitting and poor out-of-sample forecasts.

Dicks (1990) discussed the overfitting issue while estimating a number of simple demand and supply equations for the UK market based on earlier research by Hendry (1984). Dicks found that extending the demand and supply models to include the mortgage market, demographic factors, and construction costs can improve short-run forecasting results for prices and volumes and achieved reasonable results for the 1970s and 1980s, albeit with a tendency to under-predict the rate of house price increases.

Forecasting with a regression model requires conditioning the model on future values of explanatory variables. Other models will need to provide these variables, such as household income from a macroeconomic model. The forecast accuracy will be largely determined by these exogenous forecasts.

---

[24] Technically, they use 'stacked' OLS regressions, which is estimating for each city but constraining all coefficients to be the same across cities.

[25] The full specification in Auterson (2014) extends this model to an error-correction framework.

Accuracy (medium run): fair. Although forecasts using explanatory variables introduce an additional source of uncertainty, anchoring the medium run forecast to fundamentals may nonetheless provide an improvement over naïve forecasts (Lawson (2011) confirms this for the case of Californian housing prices and provides a discussion). However, as most relationships must be specified in terms of their growth rates to achieve stationary data, the medium-run performance is likely to perform poorer than other models that permit long-run levels relationships (see Subsection 3.5).

Communication (story telling): good. Multivariate regression models must be conditioned on future values of exogenous explanatory variables. This makes them well-suited for story telling and integration within a wider budgetary framework to provide a consistent budget narrative.

Communication (transparency): good. They can provide a clear explanation for forecast errors. Equations can be published and their specification (particularly model coefficients) is unlikely to change frequently. Further, model coefficients have intuitive interpretations that can be easily evaluated and repeated by budget scrutinisers with a general background in economics.

Data compatibility: fair. Multivariate regression models work well with the number of observations of quarterly data available for the Scottish housing market. The data required for the asset-price approach appears to be available, including housing starts and completions, although there may be limitations on the length of rental series. There is also sufficient data for affordability models, including interest and total payment to income ratios available from the Council of Mortgage lenders. That said, there are likely to be some restrictions on the set of explanatory variables for Scotland, rather than the UK as a whole, and the data requirements are more involved than univariate models, resulting in a *fair* score.

Resources: fair. Econometric models are not push-button and require more resources than purely statistical models such as in Subsection 3.2. They require fewer resources than other techniques in the review such as DSGE models, but nonetheless require specialised knowledge about both housing markets and econometric theory. They require frequent maintenance, re-estimation, and re-specification. Makridakis et al. (2008) provide a useful discussion of the resources devoted to econometric models versus simpler univariate approaches:

> Whether intended for policy or forecasting purposes, econometric models are considerably more difficult to develop and estimate than using alternative statistical methods. The difficulties are of two types:
>
> > 1. Technical aspects, involved in specifying the equations and estimating their parameters, and
> >
> > 2. Cost considerations, related to the amount of data needed and the computing and human resources required. (p. 302)

On the question of whether the extra burden of multivariate models over univariate approaches is justified, Makridakis et al. provide an opinion based on their own experiences, that suggests the appropriateness of a multivariate econometric model will ultimately depend on its intended use within the Scottish Government's budget production framework:

Our assessment of multivariate models is summarised in Appendix Table A3.

3.4 Vector autoregressive models

The basic vector autoregressive model (VAR) is a collection of time series models for different variables, estimated at the same time as a system. VAR models offer a simple and flexible alternative to the multivariate regression models of Subsection 3.3. The VAR approach need not rely on economic theory to specify relationships between variables (though theory often drives the choice of variables to include). VARs are instead based on the idea that economic variables tend to move together over time and tend to be autocorrelated.

Sargent and Sims (1977) promoted VARs as an alternative to large-scale macro-econometric models. They criticised macro models for the strong assumptions they imposed on the dynamic relation between macroeconomic variables and for not accounting for the forward-looking behaviour of economic agents.[26] They proposed an alternative that allows the data itself to determine how macroeconomic aggregates interact.

In VAR equations, the time path of the variable of interest can be affected by its past values and current and past values of other variables, while also letting the other variables be affected by current and past realizations of the variable of interest and each other—that is, they allow feedback between variables.[27] For a simple case of two variables, it has the following form, taken from Enders (2014):

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_{yt}$$
$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_{zt}$$

While the VAR model does not need to make any assumptions (impose restrictions) about which variables affects the other, an economic theory-based model can be imposed on a VAR, along with other behavioural and reduced-form (no contemporaneous effects) specifications.

If some series are thought to be determined exogenously or the researcher is working with ragged edge data (releases of some series are available before others) their values can be imposed exogenously. An example may be the outlook for the policy rate path of the Bank of England. However, Brooks and Tsolacos (2010) note that comparisons between unconditional and conditional VARs find little improvement in forecast accuracy from using conditioned exogenous variables. VARs can also include exogenous variables such as time trends, seasonal dummies, and other explanatory variables. Non-stationary data (as is likely to be the case for Scottish prices and transactions) may need to be transformed (using logged differences or growth rates) before entering the VAR.

---

[26] Specifically, they suggest that no variables in the economy can be taken as exogenous.

[27] This structure permits all variables to be treated as jointly endogenous (*y* is determined by *x* and *x* is determined by *y*).

Assessment

Application (forecasting): good. VARs are well-designed for forecasting and are commonly used across a wide range of forecasting applications in the public and private sector. Because they contain only lags of variables, future values can be forecast without forecasting other determinants in separate models and imposing them exogenously or assuming their time path—that is, they are not conditioned on any future realisations of explanatory variables.

Brooks and Tsolacos (1999) provide a useful example of a VAR applied to the UK housing market. They estimated the impact of macroeconomic and financial variables on the UK property market, as measured by a real estate return index. They chose monthly data to be comparable to US studies over the period December 1985 to January 1999. The variables were selected based on other non-UK studies that have determined the variables' relevancy under various theoretical and empirical specifications. The variables include: the rate of unemployment as a proxy for the real economy (as its available at the monthly frequency), nominal interest rates, and the spread between long- and short-run rates, unanticipated inflation, and the dividend yield. They find that macroeconomic factors offer little explanatory power for the UK real estate market, although the interest rate term structure and unexpected inflation have a small contemporary effect.

Even if not used as the main forecast, VARs frequently serve as a yardstick against which to measure the forecasting performance of other more resource-intensive models, such as large-scale macroeconometric models.

Application (policy): poor. It is generally difficult or impossible to recover interpretable directional causal relationships from a VAR in practice. VARs are therefore not useful for scenario analysis or fiscal impact costings. Certain structural forms of VAR models have some use for performing risk assessments, for example, the impulse response of a real income shock on housing prices.

Accuracy (short run): good. VARs are often found to perform better than univariate time series and more complex theory-based models. They are especially suited for the short-run horizon. This is conditional on having sufficient Scottish historical data that doesn't limit the VAR's specification.

Accuracy (medium run): fair. Because VARs specifications are not grounded in long-run causal relationships based on theory, forecasts for the medium-term may suffer. There were generally few applications of VARs beyond 8 quarters.

Communication (story telling): poor. By not imposing a strict theoretical structure, VARs allow the data to drive the forecast. Although this makes for a better forecast, it makes interpretation difficult. The complex lag structure (and contemporaneous impacts of variables if so specified) makes it difficult or impossible to isolate the influences of variables on each other to tell a story.

A VAR may have trouble being made consistent with other budget forecasts and the economic narrative, depending on the specification. Under certain conditions it can be constrained to other forecasts or conditioned on exogenous variables from the economic model or other fiscal variables.

Communication (transparency): fair. VAR specifications are likely to change frequently and would need to be published frequently. External budget scrutiny and testing of

24

equations would be limited to specialists. However, the limited judgment involved with running a VAR model adds to its transparency.

Data compatibility: fair. Because of the lag structure in VARs, adding additional variables increases the number of parameters that the number of parameters to estimate dramatically. More parameters require more observations. To conserve degrees of freedom, standard VARs are generally quite small, with around six to eight variables (Bernanke, Boivin, and Eliasz, 2005). Given the relative limitations of Scottish data compared to UK and US data, this is likely to be even fewer. Although the number can be expanded with Bayesian techniques (see Section 4), in practice it may be necessary to discard potentially useful variables simply to estimate the model. A problem may emerge for Scotland if there are sufficient observations to estimate the VAR, but not enough to include enough lags to whiten residuals. This issue can be surmounted by common factors analysis laid out by Bernanke et al. (2005), Stock and Watson (2002) and discussed further in Section 4.

Resources: good. VARs can be implemented quickly and largely programmatically in statistics software packages, using automatic criteria for selecting the model's lag length. They are unlikely to require significant resources or specialists.

Our assessment of VARs is summarised in Appendix Table A4.

3.5 Error-correction models

The forecasting techniques discussed so far can be used only if house prices, volumes, and explanatory variables are stationary or transformed to be stationary—that is, their means and variances are constant over time.[28] House prices and transactions in Scotland and elsewhere tend to grow over time. Further, their quarterly fluctuations (variance) tend to be different during different periods.

Differencing the series to apply ARIMA and VAR approaches allows model coefficients to be estimated with OLS regression, but may sacrifice explanatory power between variables in their levels form. Further, there would no longer be a long-run solution to the model. In economics applications, this long-run solution means that the system is in equilibrium and there are no longer short-run fluctuations.[29] This would be the case, for example, in situations where the output gap is closed and economic variables have returned to their long-run steady state (such as in the outer years of a five-year budget forecast). For the housing market, this long-run solution is generally considered as the horizon over which supply is elastic.

Error-correction models were developed to overcome the limitations of differencing to preserve the long-run levels information and present both the short-run growth information and the long-run equilibrium relation in a single statistical model. This makes them

---

[28] As discussed in Subsection 3.2, variables that trend over time with non-constant variance are called "integrated". When variables are integrated, hypothesis tests are not valid (the t-statics and F-statistics do not conform to the t- and F- distributions).

[29] Technically, if variables have returned to their long- run steady state, then the first difference:

$$\Delta y_t = \beta_1 \Delta x_t + e_t$$

does not have a solution: the changes in x and y are zero and the terms drop out.

particularly powerful for forecasting the real estate market, given that it has been shown that this long-run information contains useful information (as discussed above in Subsection 3.3).

Error-correction mechanisms are based on the concept of cointegration: two or more non-stationary variables may wander around, but will never be too far apart—that is, the gap between them is stable over time, and the gap is itself stationary.

This empirical connection is the result of a theoretical equilibrium market force or shared trend. Researchers applying error-correction models to the housing sector point to several such potential cointegrating relations:

- household incomes and house prices
- rental rates, the discount rate (interest rates) and house prices
- house prices, GDP and total employment

A basic error-correction model is represented by the following equation, based on the presentation in Enders (2014):

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2 (y_{t-1} - \gamma x_{t-1}) + u_t$$

The two difference terms (differencing is represented by Δ) are stationary. The term $(y_{t-1} - \gamma x_{t-1})$ is an algebraic manipulation of the long-run levels model, and it represents that amount by which the two variables were out of equilibrium the period before (that is, the error). For example, this could be the amount by which imputed rents from owner-occupied homes are out of sync with quality-adjusted actual rents. Because $(y_{t-1} - \gamma x_{t-1})$ is stationary, the model can be estimated with OLS and statistical inference is valid. The coefficient β_2 is the speed at which the disequilibrium is corrected (for example, a coefficient of 0.5 would mean roughly half the gap between imputed rents and quality-adjusted actual rents is closed one period later.

Error-correction models rely on the same theoretical underpinnings as the multivariate regression models in Subsection 3.3. Indeed, many of the specifications and models discussed in 3.3 are more appropriately implemented as error-correction models.

Assessment

Application (forecasting): good. There is a vast literature that applies error-correction models to the housing sector. An influential error-correction model framework for housing supply and prices was developed by Riddel (2000, 2004). It allows both disequilibrium in housing supply and house prices to affect one another. Stevenson and Young (2014) applied the model to the Irish housing market, which may serve as a useful guide for modelling the Scottish market. In this model, the long-run supply equilibrium is estimated empirically by the equation

$$\ln HC_t = \beta_0 + \beta_1 \ln HP_t + \beta_2 \ln BC_t + \beta_3 r_t + \upsilon_t$$

and the error-correction specification is

$$\Delta \ln HC_t = \beta_0 + \beta_1 \Delta \ln HP_t + \beta_2 \Delta \ln BC_t + \beta_3 \Delta r_t + \beta_4 \varepsilon_{t-1} + \beta_5 \upsilon_{t-1} + \xi t$$

where: HC_t = housing completions
HP_t = prices
BC_t = real building costs
r_t = the real after tax interest rate
$\upsilon_{t,\xi t}$ = error terms

ε_{t-1} = the lagged disequilibrium (error) from the long-run price equation below

The long-run price equilibrium is an inverted demand function, estimated by the equation

$$\ln HP_t = \beta_0 + \beta_1 \ln POP_t + \beta_2 \ln RDI_t + \beta_3 \ln HS_t + \beta_4 r_t + \varepsilon_t$$

and the error-correction specification of the price equation is

$$\Delta \ln HP_t = \beta_0 + \beta_1 \Delta \ln POP_t + \beta_2 \Delta \ln RDI_t + \beta_3 \Delta \ln HS_t + \beta_4 \Delta r_t + \beta_5 \varepsilon_{t-1} + \beta_6 \upsilon_{t-1} + \omega_t$$

where: POP_t = population aged 25 to 44
        RDI_t = real disposable income per capita
        HS_t = is the per capita housing stock
        ω_t = error term

Addison-Smyth, McQuinn, and O'Reilly (2008) modelled Irish housing supply using error-correction models and found that developers do respond to disequilibrium. However, the findings also suggest the gap is slow to correct itself, with only roughly 10 per cent of the disequilibrium being corrected annually.

Error-correction models are perhaps the most common way to model the housing market in finance ministries and central banks. For example, all three major public macro forecasters in the UK (HMT, OBR, and the Bank of England) rely on error-correction models.

Application (policy): fair. Error-correction models rely strongly on economic theory and are relevant for policy analysis. They can include a wide range of variables representing government policy and other explanatory variables that can be changed to evaluate alternative assumptions. That said, their focus is on forecasting the dynamic impact of these variables, and they rely on cointegrating relationships between variables that may not exist between the policy levers of interest for fiscal impact costing, and so are less relevant to policy than multivariate regressions or microsimulation models with a fiscal impact costing focus.

Accuracy (short run): fair. Error-correction models generally perform well in both the short run and the medium run. However, there is some evidence that in the UK they may underperform in the first eight quarters of the outlook. The OBR has found that the error-correction model it uses to model the housing market (HMT and the Bank of England use similar approaches) is not well-suited for capturing short-run dynamics, although it provides good forecasts in the medium run (Auterson, 2014).

Accuracy (medium run): good. Because of their basis in theory and being ground in long run levels equilibrium relationships, error correction models are likely to provide better forecasting performance over years three to five than other methods. Lawson (2011) found convincing evidence that error-correction models outperform a number of other univariate and multivariate models over a three-year horizon for Californian housing prices when estimated over the period 1975 to 2006 and forecast over 2007 to 2009. Larson also found that error-correction models could predict a housing price decline well in advance (the ability to forecast the timing of the decline, however, was poor).

Communication (story telling): good. Like multivariate regression models, error-correction models are conditioned on future values of exogenous explanatory variables. This makes them well-suited to story telling and integration within a wider budgetary framework to provide a consistent budget narrative.

Communication (transparency): fair. Equations can be published and their specification (particularly model coefficients) is unlikely to change frequently. Model equations are more opaque to budget scrutinizers with only a general economics background than more simple regression equations, but the equations are nonetheless more economically intuitive than VARs.

Data compatibility: fair. While it is possible to estimate an error-correction model using Scottish data, there may be some limits that could affect forecasting performance. Practitioners suggest that housing cycles modelled with error-correction models can last eight to ten years, and that these dynamics will form the basis for estimating the error-correction model's parameters. Suitable Scottish historical data may only capture one cycle, and that cycle included the financial crisis.

Resources: fair. Error-correction models are likely to require greater expertise to develop, run, and maintain, than many other options. However, the specification and forecasting can be done easily in statistical software packages by specialists, and is not likely to require significant time or effort following the initial development period.

Our assessment of error-correction models is summarised in Appendix Table A5.

3.6 Large-scale macroeconometric models

Forecasts of the housing market are produced within the macroeconomic models of budget forecasting frameworks to estimate the residential investment component of GDP, an important driver of business cycles.

Macroeconometric models simulate the economy as the interaction of aggregate supply and aggregate demand on the same basis as the National Accounts statistical framework. They use a mix of the techniques above to specify equations that describe the working of the entire economy. Although not employing new tools, the systems approach and the way it is implemented in practice—particularly national accounts data and identities and the goal of forecasting GDP—deserves special consideration as a model class on its own.

Macroeconometric models are loosely grounded in Keynes's General Theory, which Hicks (1939) and later researchers formulated into the well-known IS-LM framework. The estimation of a system of econometric equations estimated one equation at a time (ad hoc basis) related together using national accounting identities was first undertaken by Klein and Goldberger (1955) for the Cowles Commission (an initiative running from 1939 to 1955 to apply mathematical and statistical analysis to the economy).[30]

Models typically take a view of the output gap (the economy's actual output relative to its potential output, forecast separately) and combine it with other macroeconomic relationships such as the IS curve (aggregate demand and interest rates), Phillips curve (unemployment and inflation), a Taylor rule (monetary policy), and interest parity conditions (exchange rates). They strike a middle ground between theory-based and pure time-series models, taking aspects of both to capture both theoretical relationships and rich dynamics of variables over time for forecasting.

The housing sector (private sector investment in dwellings in the UK, often called residential investment or residential gross fixed capital formation elsewhere) includes

---

[30] De Vroey (2016) provides a useful history.

purchases of new housing and major improvements to existing dwellings by households (Carnot et al, 2014).[31] It is an important determinant of (and is determined by) household wealth and consumption with the model framework. It is estimated using aggregate behavioural equations within the household sector to derive the wealth stocks of consumers that, along with disposable income, guide the household sector's consumption equations.

A typical equation for the real stock of housing resembles the following combination of a short run difference and long-run levels equation (error-correction model), from Carnot et al. (2014):

$$\Delta k_t = \alpha + \Sigma_i \beta_i \Delta y_{t-i} - \mu(k_{t-1} - \eta y_{t-1} + \theta r_{t-1}) + \gamma z_t$$

where: y = real disposable income
k = residential investment
r = the real interest rate (usually a long rate, but Carnot suggests a short rate in the UK, where mortgages are predominantly at variable rates)
z = other explanatory variables such as the relative price of housing

The justification for this specification is often grounded on the neo-classical standard model of life-cycle utility maximisation, where consumers choose between a mix of consumption goods and housing investment goods. This is, however, only a loose theoretical justification—it is not necessarily implemented by specifying a household utility function, which is the realm of DSGE models discussed in Subsection 3.8. The degree of foresight and optimisation of the household can differ. They can have perfect foresight (Robidoux and Wong, 1998) or include some rule of thumb consumers, to introduce an element of constrained rational expectations (Gervais and Gosselin, 2013).

Housing prices are usually modelled as a rental price for housing services (for owner-occupied homes this is the best estimate of what the owner would charge if she were renting it to herself). The price then feeds into the rate of return on housing investment, which drives investment in residential housing and affects future supply.[32]

Traditional macroeconomic models continue to be the workhorse of macro modelling in government departments and central banks. Although DSGE models were implemented in many central banks and finance ministries, they are general used for scenario analysis in parallel with macroeconometric models and to challenge the forecast from an alternative perspective. Further, traditional large-scale macroeconometric models are experiencing a resurgence in popularity and credibility (for example, see the academic and online discussions generated by Romer (2016)).

Assessment

Application (forecasting): fair. Finance ministries and central banks are moving toward making their macroeconometric model documentation public. There are many published examples that forecasters could use as a foundation on which to build the model using Quarterly National Accounts Scotland.

---

[31] Depending on the National Accounts specification, households can also include unincorporated businesses.

[32] The investment is driven by the difference between the rental price of housing services and the costs of those funds (interest rates). This is known as a Q-ratio.

For example, the housing sector component of the macroeconometric model shared by HMT and OBR is described in OBR (2013). It forecasts private sector investment in dwellings (RES_t in £m and chained volume measures) using the following relationship with real house prices, the real interest rate, and the number of property transactions:

$$\ln RES_t = 2.07 - 0.26\ln RES_{t-1} + 0.02\ln(APH_{t-1}/PCE_{t-1}) - 0.001(RS_{t-1} - 400\Delta\ln APH_{t-1})$$
$$+ 0.08\ln(PD_{t-1} * 0.85) - 0.14\Delta\ln RES_{t-1}$$

where: APH_t = an average house price index
               PCE_t = the consumer expenditure deflator
               RS_t = UK three month inter-bank rate
               PD_t = property transactions

Property transactions (particulars delivered) are assumed to be negatively related to the difference between actual and expected house prices, where expected house prices are determined by the user cost of capital, consumer prices, and real disposable income, given by:

$$\Delta\ln PD_t = -0.11\ln PD_{t-1} + 0.25\ln RHHDI_{t-1} - 0.22\ln RHP_{t-1} - 0.002 UCH_{t-1} + 9.07\Delta\ln A2029_{t-1}$$
$$+ 0.10\Delta\ln PD_{t-1} + D_t$$

where: $RHHDI_t$ = real household disposable income
               RHP_t = real house prices, APH_t/PCE_t
               UCH_t = user cost of housing (a function of mortgage rates and the change in prices in the previous period as a proxy for the expected capital gain)
               $A2029_t$ = population of cohort aged 20-29
               D_t = a collection of dummy variables to control for abnormal events

The remaining equations of the model estimate the other components of aggregate demand: consumption (durables and current), investment, government spending, exports, and imports (see OBR (2013) for the complete specification).

There is a robust literature demonstrating the importance of the housing sector's role in the macroeconomy and importance of considering it within this wider framework

For example, the correlation between the growth in housing prices and the growth in consumption and savings has been demonstrated by Meen (2012), who estimated the correlation coefficient as 0.74 on average over 1982 to 2008. The relationship has broken down somewhat since the turn of the century, however. The correlation in the period 1982 to 1999 was 0.81 and for 2000 to 2008 was 0.65.

In the US, Case et al. (2005) provide a useful summary and theoretical framework, finding house wealth to be more important than other forms of financial wealth for driving consumption patterns.

Although the correlation between housing prices and consumption and saving behaviour is well established, conclusive evidence of causality has remained elusive. Elming and Erlmler (2016) reviewed the relevant literature and found four main links between house prices and consumption: 1) the housing wealth effect, 2) housing equity serving as collateral and precautionary wealth, 3) the common factor of income expectation, and 4) the common factor of overall credit conditions and financial liberalisation. The authors provide convincing evidence in favour of a direct causal influence of housing prices on

consumption. They do so by exploiting the natural variation between household price drops in different regions during the global financial crisis, using households with two public service income earners to control for income expectations (the public service salaries are set through strict collective bargaining arrangements).

Macro models could form part of the overall budget and help inform the housing market forecast; however, their applicability to the LBTT forecast itself may be limited. Macro models typically use a different concept of average house prices than would be useful for the tax base, and would need to be modified to serve that purpose. Many institutions use an auxiliary model, which is estimated outside the forecasting framework and they impose the model results exogenously (Auterson, 2014).

The aggregate time series in the national accounts, such as residential investments, may have limited ability to predict house transactions. Mankiw and Weil (1989) discuss how the noise (large standard errors) of national accounts data obscures relationships that show up when estimated using other time series grounded in administration or census data.

Application (policy): fair. Macroeconometric models can be used for policy analysis. For example, HMT used the macro forecast and OBR's auxiliary housing market model to assess the impact of a vote in favour of leaving the EU on house prices (HMT, 2016). They can be used to model economic and fiscal sensitivities to shocks such as an oil price decline, and prepare fiscal multiplier estimates (Office of the Parliamentary Budget Officer, 2016). However, they may be of limited use for fiscal impact costing, because of the differences between housing investment in the national accounts and the tax base.

Accuracy (short run): fair. Early macroeconomic models that were driven by theory alone, ignoring the dynamics and inter-temporal properties of the data, generally produced poor forecasts.[33]

Forecast performance has since been improved with the introduction of better techniques to capture dynamics, and modern macroeconometric models should fare relatively well for their purpose. Granziera and St-Amant (2013) compares the forecast of their rational ECM framework of the housing markets to AR and regular ECM models and finds it performs better both four quarters and eight quarters ahead in rolling forecast experiments over 2002 to 2011.

As discussed, macroeconometric models are a blunt tool for forecasting tax bases. Conforming to national accounting identities restricts the specification of the tax base (housing transactions can relate to non-current production). Correspondence between tax bases and national accounts aggregates can be poor. They tend to perform more poorly than a model dedicated to the specific tax base and tax program parameters.

Accuracy (medium run): fair. Because of their theoretical underpinnings, use of long-run equilibrium conditions (closing of the output gap), and a greater ability to maintain variables in a levels specification (making use of error correction models), these models are likely to improve upon naïve forecasts for the medium run. However, long-run macroeconometric forecasts suffer from the same concerns regarding tax base congruence as in the short run.

---

[33] Granger and Newbold (1986) provide a summary of several studies that demonstrated that univariate time-series models can outperform large-scale macreconometric models.

Communication (story telling): good. Large-scale macroeconometric models offer relatively straightforward, intuitive relationships that can be communicated easily. Parameter signs and magnitudes should make intuitive sense and align with economic theory. Because of their reduced-form specification, interactions between equations and simultaneously determined outcomes are limited relative to an unrestricted systems approach. Impulse response charts can be produced and published publicly to demonstrate the characteristics of the model.

Communication (transparency): fair. Model documentation, equations, and datasets can be published online for public scrutiny by interested parties such as academics, think tanks, and private consultancies (for example, the ITEM group at Ernst and Young uses HM Treasury's model for its own consultancy purposes, including an annual report that provides a check on the government's forecasts).[34] Parameter estimates and equation specifications should remain relatively stable between forecast. That said, the transition from history to model results involves a lot of smoothing and adjustments to fit with economic monitoring. The modeller's judgment plays a large role in the starting point, first two quarters, end point, and dynamics in between. For this reason the model only receives a *fair* score.

Data compatibility: fair. Estimating a large-scale macroeconometric model should be possible using the Scottish Quarterly National Accounts. However, drawbacks related to congruence with the tax base could be amplified as a result of Scottish data limitations, as there are only high-level sector accounts for residential gross fixed capital formation on an experimental basis.

Resources: poor. One or two experienced and skilled analysts are sufficient to maintain a large-scale macroeconometric model once developed; however, the development is a significant undertaking. Practitioners reported that it is useful to have several less-experienced analysts in charge of economic monitoring and to work independently on model components such as import/export modules. Model development would require programmable statistics software beyond Excel.

Our assessment of large-scale macroeconometric models is summarised in Appendix Table A6.

3.7 Dynamic stochastic general equilibrium models

Dynamic stochastic general equilibrium (DSGE) models are a systems framework for modelling the aggregate economy. They are relied upon heavily in modern macroeconomics. There is a broad spectrum of models falling under the DSGE label and an exact definition is difficult to pin down. De Vroey (2016) suggests that early DSGE models were defined by the following elements, based on his interpretation of Manuelli and Sargent (1987):

- a general equilibrium perspective (markets are not considered in isolation, but rather as a whole)
- dynamic analysis (the decisions and behaviour of economic agents depend are modelled over time, rather than a single period)

---

[34] See the ITEM team website at: http://www.ey.com/uk/en/issues/business-environment/financial-markets-and-economy/economic-outlook.

- rational expectations (the forecasts of agents within the model are assumed to be the same as the forecasts of the model (see Sargent, 1987))
- microfoundations (economic agents are modelled with optimizing behaviour, for example utility-maximizing and forward-looking behaviour in household decisions on saving, consumption, and labour supply subject to a budget constraint)
- markets clear (prices adjust to eliminate excess demand or supply)
- exogenous stochastic shocks (shocks come from outside the system rather than emerging within)

These features, along with other elements such as production technologies, budget constraints, and decision rules are formulated mathematically to represent the economy in a manner that a computer can simulate and solve. Equations are estimated all at once in a binding, unified way rather than the piecemeal equation-by-equation method of traditional macroeconometric models (Carnot et al, 2011).

Although early models assumed macroeconomic fluctuations were the result of random disturbances in technology and preferences, they were eventually modified to be based on frictions (price and wage rigidities) and to include monetary policy, with these new classes of models being called New Keynesian models (for example, see Christiano, Eichenbaum, and Evans (2005)).

DSGE models emerged out of the real business cycle literature, notably Kydland and Prescott (1982). Slanicay (2014) provides a history, describing them as a response to the forecasting failure and problematic theoretical underpinning of large-scale macroeconometric models—namely, the simultaneous high inflation and high unemployment of the 1970s (a breakdown in the Keynesian Phillips curve) and lack of microfoundations.

A technical specification of the equations of DSGE modelling would go beyond what is possible in this review, but we provide a qualitative description of a small-scale new-Keynesian DSGE model, based loosely on An and Schorfheide (2007) as presented by Herbst and Schorfheide (2015).

The basic DSGE economy is modelled using five agents. There are two types of firms: a single representative **final goods producing firm** that is perfectly competitive, taking input prices and output prices as given. The firm's inputs are supplied by **intermediate goods producing firms** that are monopolistically competitive, choosing labour inputs and output prices to maximise the present value of future profits. **Households** choose their labour supply, consumption, and saving to maximise utility, subject to a budget constraint that accounts for investment returns and tax payments. A **central bank** uses an interest rate feedback rule to respond to monetary policy shocks and targets a steady-state inflation rate consistent with a level of output at its potential. A **fiscal authority** is assumed to consumes a fraction of aggregate output subject to a budget constraint and levies a lump-sum tax to finance any shortfalls in government revenues.

Examples of DSGE models that could be drawn upon for Scottish forecasting include the Bank of England's COMPASS, the Bank of Canada's ToTEM, and the New York Federal Reserve's FRBNY models. The IMF also uses two well-known DSGE models: MultiMod and GEM.

DSGE models have faced criticism following the financial crisis and their use and misuse is being keenly debated. Their opponents include Romer (2016), who criticises DSGE

models broadly, attributing their popularity to "imaginary" constructs that succeeded through mutual loyalty among well-known economists and a departure from scientific principles.

A particular target of criticism is the model's rational expectations assumption, although methods to introduce frictions have been implemented to slow adjustments to reflect observed behaviour, and more recently models such as Slobodyan and Wouters (2012) have introduced bounded rationality.

Assessment

Application (forecasting): poor. The focus of DSGE models is not forecasting, but rather simulating and tracking how shocks are propagated through the economy. Slanicay (2014) describes their range of application, from the models that central banks use to discuss the transmission of monetary policy shocks through the economy, to the more stylised academic models tailored to test and demonstrate implications of particular economic assumptions.

Applications of DSGE models to the housing market are limited. Basic DSGE models typically only operate in flow space (changes period-to-period) and residential investment stocks generally do not play a role. That said, there have been some efforts in recent years to incorporate a housing sector.

For example, Caldara, Harrison, and Lipinska (2012) developed a method to use the correlations between housing price shocks estimated in auxiliary VAR models with variables included in the DSGE model to assess the implications of shocks to US housing market data.

Other research comes from Iacoviello and Neri (2010). The authors considered both the impact of macroeconomic shocks on the housing market and how shocks to the housing market affect the macroeconomy. The housing market is incorporated via a production function that produces houses using capital, labour, and land.

A recent innovation in DSGE models is the stock-flow consistent models of Burgess, Burrows, Godin, Kinsella, and Millard (2016). They show promising improvements on traditional DSGE models, incorporating the balance sheets of economic sectors including the housing sector. Stock-flow consistent DSGE models may be more suitable for housing market forecasting and policy in the future.

Application (policy): fair. DSGE models may have use for modelling the transmission through the economy of housing market scenarios. DSGE models are not appropriate for static fiscal impact costing (costings that are estimated using a single market and do not consider the feedback effects of the rest of the economy). However, they can be used for dynamic scoring (modelling and costing the feedback of government policy changes through the wider economy).

Accuracy (short run): fair. Although early DSGE models had poor forecasting performance, recent developments such as Smets and Wouters (2007) have demonstrated refinements that can offer better forecast properties. The *fair* score has been given for the general forecasting performance of DSGE models, but forecasters should consider that application to the housing market is largely untested.

Accuracy (medium run): fair. DSGE models may offer better performance in the medium run than other models less grounded in economic theory. Iacoviello and Neri (2010) find

their housing-market augmented DSGE model is able to capture long-run trends and the medium-run business cycle well. When choosing a DSGE model for the medium-run horizon, there is evidence that smaller is better. For example, Del Negro and Schorfheide (2012) assess the forecast performance of the large-scale Smets and Wouters (2003) DSGE model against a small-scale version. They find that although the short-run forecast performance of the large-scale model is slightly better than the smaller model, the medium-term forecast performance of the compact model is superior.

Communication (story telling): fair. DSGE micro foundations lend themselves to intuitive narratives. However, complexity of interactions runs significant risks of becoming a 'black box' with difficult or no interpretation.

Communication (transparency): poor. Caldara et al. (2012) suggest that as layers of complexity and interaction are added, the results and interactions become more opaque and harder to explain to policymakers. Considerable judgment is applied throughout estimation. External scrutiny would require specialist training.

Data compatibility: good. A DSGE model could be estimated using the Quarterly National Accounts of Scotland. The Smets and Wouters (2007) model is constructed using only seven data series (real GDP, consumption, investment, wages, hours worked, inflation, and interest rates). To model the housing sector in the manner of Iacoviello and Neri (2010) would require the addition of measures for capital and land.

Resources: poor. DSGE models are generally an advanced forecasting technique requiring specialised training and most likely the services of a PhD economist, especially during development. However, the required resources may not put DSGE models out of reach of the Scottish forecasting framework.

Our assessment of DSGE models is summarised in Appendix Table A7.

3.8 Microsimulation models and micro databases

Microsimulation models use survey data to construct a representative distribution of typical households and individuals in the economy.[35] Weights are then used to scale the sample to the population level. If linked to tax returns, microsimulation models can be used to assess the fiscal and distributional consequences of changes to the tax and transfer system.

Microsimulation models are not designed for forecasting—they are static accounting models that mechanically apply legislated or proposed tax and transfer parameters to the relevant characteristics of individuals and households. These may include characteristics such as an individual's net income, the number of children under a certain age that qualify for child benefits, or real estate purchases made throughout the year.

Certain properties of microsimulation models nonetheless allow them to be used as a tool in a wider framework to arrive at forecasts. Specifically, models can apply the current tax system to the population characteristics in past years. In this manner, forecasters can calculate the revenue elasticity (sensitivity of growth) of a tax to the tax base. For a flat tax, this revenue elasticity would be zero. However, for a graduated tax such as personal

---

[35] As most survey data focuses on individuals, there are few microsimulation models for businesses; however, some simple micro data models make use of Corporate Income Tax returns.

income taxes or LBTT, revenues will increase faster than one-to-one with the base.[36] This is known as *fiscal drag*. Because LBTT thresholds are not automatically indexed to inflation, fiscal drag could be significant.

The elasticity of revenues to the tax base can then be applied to a forecast of the tax base to arrive at a forecast of revenues. Although the model may be built with the ability to grow the base itself to future years, the benefit of this approach is that the forecast doesn't have to correspond exactly to the tax base of revenues. For example, the microsimulation-based elasticity can be imposed on a macroeconometric model by using its historical correlation with the National Accounts elasticity.

Although not technically a microsimulation model, some practitioners use a database of the universe of tax returns that can be queried to retrieve variables of interest. These can include simple relationships that can be programmatically changed and aggregated to cost alternative policies, or can be assessed before and after a policy has been implemented to assess its impact (for example, see Matthews (2016)).

Microsimulation models and micro databases are also among the few ways to examine the distribution of housing prices over time. If clear trends are observed in the summary statistics of the distribution (for example, its mean, variance, and symmetry), these may be forecast using techniques described above such as univariate time series models.

Assessment

Application (forecasting): poor. Microsimulation models and micro databases cannot, on their own, provide forecasts for the housing market. However, they can be used in conjunction with other forecast models to incorporate fiscal drag effects into the forecast. For example, their base years can be grown to future years by imposing growth rates from auxiliary forecast models, and tax rates and thresholds can then be applied to this uplifted base to project tax receipts.

Microsimulation models can be used to estimate variables to include in other models, such as multivariate econometric models. For example, microsimulation models can be used to calculate the average effective tax rates of homebuyers over history and the outlook to use as an explanatory variable when estimating the net benefits of home ownership.

Application (policy): good. Microsimulation models and micro databases are particularly well-suited for policy analysis. They are the main way that government budget authorities translate forecasts of the tax base into forecasts of revenues.

There are many examples of microsimulation models in the UK applied to policy analysis. HMRC has a microsimulation model specifically for SDLT that covers the universe of transactions. Other microsimulation models in the UK include the Department of Work and Pensions' Pensim2 model, the IFS's TAXBEN model, and the London School of Economics' SAGE model. However, these models don't have a detailed specification for SDLT or Scottish LBTT.

The OBR uses HMRC's residential stamp duty plus model (SDM+) to prepare its devolved forecasts for Scottish residential LBTT, Stamp Duty Land Tax, and Welsh residential SDLT. OBR (2016) describes the SDM+ model as follows:

---

[36] This will be the case for personal income taxes even if tax thresholds are grown with inflation as a result of real income growth.

> [SDM+] allows us to apply the tax schedules for LBTT and SDLT to a full sample of transactions from a given year and then grow them in line with our price and transactions forecasts for the residential property markets (p. 22).

Microsimulation models are general developed for social policy analysis, personal income taxes, and consumption taxes. There are few microsimulation models applied to housing.[37]

Many studies assessing the behavioural response of UK policy measures simply use micro data files with the universe of tax returns to evaluate historical policy changes by looking at the period before and after the intervention. For example, Best and Klevin (2013) used the universe of SDLT administration data (that is, every property transaction) from 2004 to 2012 to assess a number of tax changes, including the success of the stamp duty holiday in 2008 to 2009 as fiscal stimulus.[38]

Microsimulation models are also useful for producing revenue sensitivity estimates, for example in the production of cyclically adjusted budget balanced and fan charts.

One drawback is that, because they are by nature mechanical, they require ad hoc adjustments to the simulation output to incorporate behavioural responses.

Accuracy. As the model itself cannot forecast, this criterion is not applicable. However, if economic variables forecast with auxiliary models are accurate, then the microsimulation model should give accurate conditional estimates of revenues.

Communication (story telling): good. Because microsimulation models are a rote imposition of the tax code on real households, communication of microsimulation results are intuitive.

Communication (transparency): good. The underlying equations are mechanical identities, and aside from weights to scale results to the population level, little estimation and no judgment is applied. Although model code can be a challenge to access and examine, results can typically be tested against back-of-the-envelope calculations.

Data compatibility: poor. There may be limitations to Scottish forecasters' access to tax-payer level data. Microsimulation model development may therefore only be possible if appropriate data protocols can be put in place.

Resources: poor. Microsimulation models require considerable resources to develop and maintain. The initial design requires a great deal of time and expertise on both tax policy and software development. However, the model may not need to be built from scratch. Li and O'Donoghue (2013) surveyed microsimulation models and point to four generic software programs that can be adapted to build new microsimulation models: ModGen (Wolfson and Rowe, 1998), UMDBS (Sauerbier, 2002), GENESIS (Edwards, 2004) and LIAM (O'Donoghue, Lennon, and Hynes, 2009).

---

[37] Exceptions include HMRC's SDM+, SustainCity (Morand, Toulemon, Pennec, Baggio, and Billari, 2010). and Australia's DYNAMOD I & II.

[38] The authors find that the stimulus was "enormously" successful, increasing transactions volumes by as much as 20 per cent in the short run. This magnitude was partly the effect of moving transactions forward, but persistent (though smaller) effects were also observed over the medium term following the end of the policy.

Parameters need to be updated to reflect new tax and transfer legislation once or twice annually within the budget cycle. This can be an involved, resource-intensive process if policies are implemented that aren't simple rate or threshold changes.

Our assessment of the qualities of microsimulation models in relation to the Scottish budget forecast process is summarised in Appendix Table A8.

4. Extensions and complements

The following techniques and approaches could be applied to refine or augment the models described above to address methodological shortcomings or to offer researchers an alternative perspective.

Bayesian methods

Bayesian techniques have been applied across the different model classes evaluated in Section 3. Bayesian inference allows forecasters to use information not contained in the history of the variables of interest (the estimation sample). This is done by constructing prior distributions that describe the state of knowledge about parameters before observing the history of the outturn variables of interest— that is, the forecaster treats parameters that are known (or roughly known) as given (that is, they are not estimated within the model), and estimate the remaining unknown parameters.[39]

Bayesian inference can be useful under conditions of limited data to develop more complex, larger models, with more variables and more parameters. This may be relevant for the Scottish market, where data may be limited compared to the UK as a whole and other markets such as the United States.

Bayesian techniques are commonly applied to the VAR class of models discussed in Subsection 3.4. A Bayesian VAR (BVAR) is a restricted VAR, in the sense that an informed value of one or more parameters is imposed on the model for estimation, rather than it being estimated over the historical sample. Specifically, instead of eliminating lags to cut down on the number of parameters that need to be estimated, the BVAR approach imposes restrictions on the coefficients of long lags by assuming that they are more likely to be zero than coefficients with shorter lags.[40] Another common restriction is that the coefficient of the first lag is assumed to have mean equal to one (Litterman, 1981).

The performance of BVARs for the housing market was assessed by Das, Gupta, and Kabundi (2010). They investigated whether information in a macroeconomic dataset of 126 time series can be used by large-scale BVAR models to forecast real house price growth in the US. They compared the forecast to small-scale VAR models, finding that forecasts over the first one- to 12- months ahead do not outperform smaller VARs.

Dynamic factor modelling

To overcome some important drawbacks of VAR modelling, such as the degrees of freedom problem (described above in Subsection 3.4), researchers have proposed

---

[39] This contrasts with the traditional frequentist approach of estimating all parameters from historical relationships. Gelman, Carlin, Stern, and Rubin (2004) refer to the traditional approach as retrospective evaluation.

[40] This is done, for example, by setting normal prior distributions with zero means and relative small standard deviations as the lags increase.

incorporating vast data sets in a procedure called dynamic factor modelling (similar—and often synonymous—concepts include principal-components regression and factor-augmented forecasting). Dynamic factor modelling uses both observed and unobserved influences to forecast a system. To do so, the forecaster first examines the variables that can be observed and their relation to other observed variables. By subtracting this shared influence from the total behaviour of the system, all the unobserved factors influencing outcomes can be backed out. The behaviour of all the unobserved factors considered together can then be used to estimate a relatively compact equation.

Dynamic factor modelling can offer the best of both worlds—a model conditioned on a very rich, expansive data set, with the statistical advantages of working with only a small number of series. It relies crucially on the assumption that a large number of time series can be summarized by a small number of indexes.

Dynamic factor modelling has been applied to the housing market by researchers such as Gupta, Jurgilas, and Kabundi (2010). They used the factor-augmented VAR approach of Bernanke et al. (2005) and a large data set of 246 quarterly series over 1980 to 2006 to assess the impact of monetary policy on real housing price growth in South Africa. Gupta et al. found that real price growth responds negatively to interest rate increases, and that the effect is considerably larger for samples of large houses than small.

While some software packages are beginning to introduce libraries and add-ons for dynamic factor analysis, there would still be significant technical challenges to overcome, requiring analysts with a very specialised skill set.

Computable general equilibrium models

Computable general equilibrium models are similar in certain assumptions to DSGE models; however, CGE models are a snapshot in time, based purely on economic theory, and do not bring in the statistical dynamics of time series across time. Their evaluation algorithms provide a static long-run steady state solution for a set of prices that equates aggregate demand for all commodities and factor inputs with aggregate supply. They are not generally suited or used for forecasting.

That said, Dixon and Rimmer (2009) proposed a framework for CGE models that could be integrated into a macroeconomic forecasting framework to produce forecasts for the household sector. They run historical simulations for each year over history. By using that information and constraining the model to a separate forecast of macroeconomic variables, they arrive at forecasts for disaggregated individual commodities and sectors such as housing.

CGE models are useful for policy applications, most suitably tax efficiency and incidence analysis, international trade, and environmental analysis. HMRC uses a CGE model to assess these issues as well as the impact of tax changes on the macroeconomy.[41]

Maintaining a CGE model would require a dedicated team, and someone with a very specific skillset (almost certainly a PhD economist with research in that field). The cost of these external consultants can run between one and two full-time equivalent employees.

---

[41] HMRC's model is programmed in the General Algebraic Modelling System (GAMS) software using the mathematical programming system for general equilibrium (MPSGE) solver.

Predicting turning points

A useful trait of a new modelling approach would be the ability to predict market corrections. This is important not only for LBTT revenues, but also for the macroeconomic outlook and a wide range of policy reasons, given the power of the housing market to influence the rest of the economy.

While we came across few methods that were demonstrated to reliably predict when future bubbles will form and burst, *probit* models hold some promise to recognising whether the current period is a peak or trough.[42] Probit models estimate a result that lies between zero and one that reflects the probability of an event occurring, in this case the probability that a given quarter is a peak or trough in real house prices (with values lying closer to one meaning the event is more probable).

Rousová and Noord (2011) estimated a probit model across 20 OECD countries to predict possible peaks and troughs in real house prices in 2011 and 2012.

Estimation of the Rousová and Noord model involves three steps:

1. Identify local peaks and troughs over history using an indicator function that identifies local maximum and minimum over an arbitrary interval (the authors use 6 quarters, following Girouard et al. (2006) and Van den Noord (2006), but others, such as Crespo Cuaresma (2010) use as few as two).

2. Impose thresholds of minimum price changes before or after the turning points identified in Step 1 to screen out minor fluctuations, leaving only major peaks and troughs. The authors use an average increase of 15% during upturns (trough to peak) and 7.5% for downturns (peak to trough).

3. Estimate two probit models to calculate the probability of being in a peak or trough in a given quarter.

$$P_t = Prob(Peak_t = 1|X) = \emptyset(\beta_0 + \beta_1 x_{1t} + \cdots + \beta_p x_{pt}$$

$$T_t = Prob(Peak_t = 1|Z) = \emptyset(\beta_0 + \beta_1 z_{1t} + \cdots + \beta_q z_{qt}$$

where P_t and T_t are the probability of a peak or trough and X_t and Z_t are explanatory variables which include the gap between actual prices and the time trend, the percentage change in real house prices, the number of house prices peaks in other OECD countries, borrowing conditions (the long-run level and percentage change in interest rates), and others.

Their model could classify 63 per cent of observations correctly. The power was higher for peaks (70 per cent). There were only two troughs in the 20 countries, which the model did not pick up (although the probability rose) but the model correctly classified when troughs were not occurring 91 per cent of the time.

Forecasting tax receipts directly

Rather than forecast the tax base and apply a complex tax model on top of the base, it is worth examining the forecast properties of several of the above techniques applied to

---

[42] In addition to probit models, Larson (2011) found that error correction models of housing prices in California before the financial crisis could forecast large declines before they were observed (although the ability to predict the timing was poor).

LBTT receipts themselves. Technical assumption-based forecasts, univariate time series, or simple structural approaches are best suited for forecasting tax receipts directly.

Fewer resources and modelling effort are likely to be required, which may be appropriate given that LBTT is small relative to the Scottish budget. Practitioners in smaller forecasting groups mentioned that for forecasts below a threshold (such as £3 billion), they simply grow tax receipts with population and inflation, or directly in-line with nominal GDP, with no estimated parameters. Given that thresholds for LBTT are not indexed to inflation, it would be reasonable to augment this with an elasticity of revenue growth to GDP to account for fiscal drag which could be estimated over history.

If not used for the model itself, a simple growth model of receipts can be a useful check on the reasonableness of the main forecast.

5. Suite modelling, combining forecasts, and averaging forecasts

As discussed in Box 1, it is unlikely that any one model can fulfil both the forecasting and policy requirements of public budget forecasters. Even if the model is intended for forecasting alone, the literature on forecast performance in Section 3 and the techniques to improve upon standard forecasts in Section 4 suggest that it is unlikely that one forecasting approach will suit all circumstances.

In response to these issues, many practitioners employ several models to improve forecasting performance or provide additional detail for policy development. Multiple models are general employed in three main manners: suite modelling, forecast combinations, and forecast averages.

Suite modelling. The demands placed on a housing market forecast could require a broader macroeconomic forecasting framework and extensive use of auxiliary models. If an objective of the housing market forecast is to include influences of the wider economy, such as employment and real income growth, then out-of-sample forecasting may require exogenous forecasts of these variables. Further, these variables could in turn depend on the housing market, and will need to be a part of an iterative macroeconomic forecasting framework. Not having this broader framework would restrict the modelling approaches that could be used.

Macroeconometric models may lack the detail required to model certain tax bases, or national accounting concepts may not be directly comparable with tax bases. In this case, multivariate econometric models estimated on their own can be combined with large-scale aggregate macroeconometric models to provide the detail necessary for bottom-up fiscal forecasting.

A suite approach may be particularly important in turbulent periods, when the economy is subjected to significant structural changes. OBR (2013) describes their approach to modelling during times of uncertainty:

> […] it makes sense to use a range of approaches and models to inform the forecast, rather than to rely solely on the behavioural relationship implied by the model, which will reflect the behaviour of the economy over the specific time period over which the relationships were estimated. (p. 3)

The Bank of England makes extensive use of suite modelling, and applies several dozen models to many of its monetary policy decision frameworks. According to Burgess et al. (2013), the Bank uses 15 or more statistical forecasting models running from univariate time series to Bayesian and factor-augmented VARs to produce a wide range of results for GDP and inflation.

In their words:

> These more traditional models have undoubted strengths: they are usually simple to understand, can quickly identify potential inconsistencies in COMPASS-based forecasts, and in many cases have an established track record in the Bank's forecast process. However, they also have limitations when compared with more structural models. They are not designed to produce joint forecast densities for the complete set of COMPASS observable variables, which makes direct comparison problematic. Moreover, in some cases, they can only produce conditional forecasts, taking some variables from COMPASS and other suite models as inputs. As a result, their forecasts may not be fully independent of all the judgements captured in the central organising model. (p. 47)

The typical use of a suite of models to ensure new policy measures are reflected in the fiscal and macroeconomic outlook was described in a practitioner interview as follows (using a change in mortgage insurance requirements as an example): 1) estimate the direct impact on GDP of changes in new and used home purchases using a micro-simulation approach based on borrower-level data from the public mortgage insurance provider, 2) estimate the broader economic implications by imposing the direct GDP shock onto the finance department's macroeconometric model, 3) update the fiscal model with both the detailed housing sector data and the new set of economic determinants from the macroeconomic model, and finally 4) iterate the fiscal model results between the macroeconometric model and fiscal model until a steady-state solution is achieved.

Simple univariate models that only consider seasonality are frequently used by practitioners to build up in-year estimates from monthly receipts. For example, if LBTT receipts exhibit a predictable monthly profile, it could be exploited to arrive at an estimate for the current year that is likely to be better than a forecasting model's raw output, depending on how many months are available. This profiling can be adjusted based on expected economic developments using the series' historically estimated sensitivity to the economy and any abnormal transactions can be removed. This careful monthly monitoring may be particularly important for Scotland's non-residential market, where a small number of high-value commercial property transactions can cause a significant spike in a month's receipts and may be carried through to future months by the model.[43]

Combining forecasting approaches. Aspects of the different model classes assessed in Section 3 can be combined—either directly in the model's specification, or by relying on different modelling approaches for different time periods.

For example, error-correction model techniques are often applied in a systems VAR approach to give vector error-correction models. Large-scale macroeconometric models can be combined with DSGE modelling approaches to create a more compact hybrid approach, as in OBR (2012b).

---

[43] See the discussion on 2015-16 non-residential LBTT revenues in the Draft Budget 2017-18 Devolved Taxes Methodology Report, available at http://www.gov.scot/Publications/2016/12/6669.

Forecasting models could be joined to forecast different time horizons. For example, a model that excels at the near-term could be combined with a model that excels at the medium- to longer-term. Or the end of the forecast horizon could be anchored with a technical assumption and then a model result used for the near-term, interpolating between. The Scottish Government previously used this approach for prices, using an ARIMA model for the first two years of the outlook and then anchoring the end of the forecast with the historical average growth rate of prices.

Combinations could be used to fill data gaps, such as the combination of a private sector average for the first two years followed by a rule of thumb (prices grow in line with average earnings) for the third year to the fifth, as the OBR used before introducing their own model. The OBR continues to use a combination approach, using their auxiliary error-correction model of the housing sector for outer years and leading indicators and judgment for earlier quarters of the outlook (Auterson, 2014).

Carnot (2014) suggests that the econometric approach to forecasting the housing sector is generally less useful during the first year of a forecast than monitoring building permits and other high frequency data for residential construction.

Averaging forecasts. There is considerable evidence that a combination of forecasts can often have better forecasts than any one approach on its own. For example, An de Meulen, Schmidt, and Micheli (2011) find that combining an AR, multivariate regression, and VAR substantially reduces the MSFE for forecast of a German house price dataset. More generally, Granziera et al. (2013) look at combining forecasts for a wide range of macroeconomic variables and find it generally improves forecasts relative to a benchmark.

Averaging internal forecasts is not done formally among any practitioners with whom we spoke. However, the same effect is achieved via judgment, challenges from other internal models, or comparisons to other non-government sector forecasts.

6. Practitioner comments

This section contains general advice and guidance we received from practitioners in governments, central banks, and private sector forecast groups that may be useful in choosing a forecasting model and creating a forecasting workflow.

The most commonly used modelling approach is a suite of models that includes a reduced-form macroeconomic model and smaller auxiliary models to provide greater detail of tax bases. In the UK, institutions rely most commonly on error-correction models for the housing tax base. The results of the auxiliary models are taken exogenously into the macroeconomic model (the macro model is told to ignore its own housing equations and take the auxiliary results in their place).

The outputs of the economic model are then sent to policy teams for estimating the cost of new tax measures and to fiscal forecasting teams for updating fiscal forecasts. Forecasting teams produce a pre-policy baseline to which costings are imposed to arrive at a post-measures outlook. Fiscal forecasts are then sent back to the macroeconomic modelling team for iteration. There are often challenge meetings between fiscal forecasters and economic forecasters at this stage to ensure that revenue forecasts are consistent with macroeconomic developments.

Microsimulation models and micro databases are the main way that government policy teams translate forecasts of the housing tax base into forecasts of revenues.

Roughly half of the forecasters had a DSGE model that was used for simulation and as a challenge to the other modelling results. Many government departments had specialty research units that provided in-depth housing research that was used to challenge the forecasts of the main models. Practitioners suggested that DSGE modelling is less relevant for forecasting, especially the housing tax base for LBTT. They suggested it could, however, be used for analysis such as examining the impulse response of an LBTT holiday on household decisions and economic activity.

All practitioners we interviewed emphasized the role of incorporating expert judgment in the forecast. This was said to be particularly important for the first two quarters to help smooth the transition from historical data to the pure model result, or to incorporate monitoring data (more recent releases and peripheral data releases that may be on a higher frequency than the quarterly forecast data). Often national accounts data conflicts with tax data, and given the former may tend to be revised and the latter are more firmly grounded in actuals, recent national accounts quarters may be adjusted. Judgment also enters during the challenge meetings, where the narrative between macro forecasters is squared with the narrative from the tax forecasters. Judgment was also applied to bring extreme forecasts closer to market or consensus expectations.

Expert judgment tended to refer to the judgment of those with domain knowledge of economic forecasting and tax administration. While we were not informed of any real estate specialists in national government forecasting units, they were included on the housing teams of private sector bank economics departments and a subnational government. The latter expressed the importance of an industry expert with deep knowledge of the real estate market. The same subnational government had legal professionals formerly in the real estate industry whose expertise was valued. Most practitioners did, however, engage regularly with the real estate business community.

One government modelling group valued its close relationship with the public mortgage insurance provider, allowing for much richer borrower-level data for policy analysis such as modelling changes to mortgage rules.

Team sizes in forecasting departments responsible for modelling the housing sector vary between one and six analysts (one senior and the rest juniors), with roles shared between housing research and other macroeconomic forecasting and special issues analysis. Typically, different teams are responsible for forecasting the housing market tax base than teams responsible for applying the tax structure to the base.

Practitioners were divided roughly evenly between three schedules for updating a model's data: 1) quarterly, 2) as soon as data is released, and 3) during the forecasting rounds for document production (for example, finance departments often update data during the run-up to spring and fall budget statements).

Parameters are re-estimated generally once a year (for example, with the release of the Blue Book in the UK) to every two years. Re-estimation is rarely left longer than two years; however, the updates depend on analyst turnover and expertise. Some teams let the data dictate—any time there is a substantial revision in the accounting procedures of the economic accounts, for example. Extensive model evaluation and development is carried

out infrequently. Some departments have internal or external evaluation committees that review the forecast every three to 10 years.

HMT and the OBR's macro model uses the WinSolve software suite, provided for free online.[44] Other practitioners listed most commonly Eviews (for its ease of handling time series data), followed by Stata. Matlab is a popular choice for DSGE modelling and Dynamic Factor Modelling. Departments with access to taxpayer data use either Microsoft Access or SQL databases for retrieving records for analysis. Microsimulation models are built as proprietary standalone software applications. Spreadsheets are used by all practitioners, particularly for auxiliary models.

Models most commonly used data at the quarterly frequency. Practitioners reported that quarterly modelling is able to capture seasonality and behavioural responses to policy changes that could be lost in annual data. Monthly data could be considered for models that require more observations, but is available for only a subset of variables. Practitioners using quarterly models often built up short-run forecasts using monthly data such as RICS, Nationwide, and Halifax indexes. Practitioners flagged the problem of identifying appropriate price indexes to use as the average house price for the dependent variables in modelling, with several different options available, each with different strengths and weaknesses.

Practitioners were split on the importance of forecast assessments and error decompositions. Roughly half do both, while the other half do not perform forecast assessments. Some suggested that accuracy in outer years is not a high concern and assessments are problematic, even if undertaken. Problems with forecast error evaluation include: changes in accounting methods (national accounts and public accounts (IFRS or GAAP accruals standards for tax revenues, for example); difficulty controlling for policy initiatives, which are themselves forecasts with large uncertainty (and often with no history to rely upon); and uncertainty and errors introduced from the forecasts of exogenous variables (for example, economic determinants such as household incomes and employment). One example of the difficulty of forecast evaluation was that a forecast's accuracy and influence itself could lead to forecast errors—for example, if a weakening housing market is forecast, the government may implement fiscal stimulus (such as the Stamp Duty Land Tax holiday following the financial crisis). The impact of these influences on the forecast error is difficult or impossible to calculate, although some try.[45]

7. Summary and going forward

Table 3 provides a summary of the individual model assessments for comparison. The weight assigned to each criterion for model selection will depend on the forecaster's priorities. We provide here a discussion of the broad themes in the literature review and practitioner interviews.

Most of the literature focused on housing prices, with decidedly less attention paid to detailed models for transactions. Little attention is paid to the distribution of prices. Many

---

[44] http://timberlake.pt/software/winsolve/

[45] For example, this forecast error decomposition plays an important role in HMRC's Measuring tax gaps reports (various). The 2016 report is available at:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/561312/HMRC-measuring-tax-gaps-2016.pdf.

academic researchers dealt with returns to real estate investment trusts and other indexes that look at investment returns, rather than forecasting average house prices themselves.

The models that were most popular and relevant for forecasting house prices included error-correction models. Particularly, the most influential literature on practitioners in the UK was the work of Geoff Meen (see Section 8: References).

The models that the literature and practitioners suggested would be less relevant are large-scale macroeconometric forecasting models and DSGE models, although these may be suited for policy analysis and other components of budget production.

On identifying turning points and structural breaks, there seem to be few clear forecast solutions, although probit models may contain useful information for whether the current period is a peak or trough, and error-correction models may have some ability to predict large future price declines. Practitioners suggested that identifying a future turning point largely relies on analyst knowledge, experience, and insight.

Studies rarely pointed to clear winners of modelling techniques for accuracy, and the consensus is that conclusions are specific to the region and time period under consideration. For this reason, many practitioners use a suite of models (typically an auxiliary error-correction model that helps inform or is imposed on a macroeconomic model).[46] The appropriate models will depend on the resources, technical expertise, and protocol of the Scottish forecasting framework.

Even if a combination of models is not used for the forecast, it can be useful to have a simple and reliable benchmark against which to evaluate alternative forecasts, as emphasized by Stock (2002), among others. Evaluating forecast errors on their own will not indicate the success or failure of a model, unless it can be compared against another feasible alternative.

On communication, the review suggests multivariate regression models are preferred to univariate models, and small systems are preferred to large and complex macro and DSGE models. The IMF's Manual of Fiscal Transparency provides useful guidance on best-practice budget forecasting, emphasizing the importance of linking revenue forecasts to macroeconomic variables, rather than simple trends or autocorrelation approaches.[47]

Practitioners almost universally use quarterly data; however, many UK peer-reviewed academic studies use monthly data. The monthly data was chosen largely for comparability with US studies and researchers admit that it restricts analysis to a smaller subset of explanatory variables.

No forecasting approaches are ruled out by data available in Scotland; however the performance of some models may be limited by a relatively short history for estimation. Further, gaps could be identified and improved in the future. To expand available data, a logical avenue to explore would be data for the UK as a whole. However, before the devolution of LBTT, the OBR found that the share of UK SDLT from Scotland varied considerably, averaging between 3.8 to 6.7 per cent (OBR, 2012a). They also found that the average property price in Scotland was below the UK average. Generally, researchers

---

[46] The OBR uses the term "imposed" instead of "exogenous", making the distinction that the forecasts have been determined outside of the model, but using the model's solutions, so are endogenous but done off-model for practical reasons.

[47] Available at: http://www.imf.org/external/np/fad/trans/manual.htm.

find that UK-wide data tends to be driven by the London housing market, where prices are likely to be influenced by different factors than prices in Scotland.

Throughout the analysis, it was taken for granted that housing prices were easily defined. There are many alternatives formulas to compute a price index, which vary in their statistical properties, such as the influence of extreme values. Forecasters could experiment with different indexes to see which best translate into revenues.

Although LBTT is a small revenue source, the housing market plays an important role in the economy. Forecasters may wish to estimate future housing demand, rents, and prices for reasons other than revenue, such as macroprudential stability, real estate market stakeholder engagement, or implementing and monitoring social housing policy. Further, the importance of a detailed treatment of balance sheet items such as household capital formation in macro models is being increasingly recognised (for example, see Mian, Rao, and Sufi (2013)). Forecasters may therefore wish to invest significant resources on that basis, and the tax revenue forecast will come out of that process at little marginal cost. If housing transactions and prices are desired only for the revenue forecast, then a simple assumption such as the private sector average may reflect an appropriate share of resources.

Next steps

A broad conclusion from our review is that there are no clear winners of forecast models for the purposes of public budgets. Model selection requires budget officials to set priorities for the model's use (the balance between forecasting and policy analysis) and the forecast's role in the wider budget (particularly in the economic outlook).

Once the model's requirements have been established, model selection will largely be determined by the specific data and circumstances of the Scottish housing market. It is difficult to determine in advance which model will perform best without developing and testing them against one another using out-of-sample forecasts to discover their comparative forecasting properties and relative practical merits. Even if the appropriate model is clear, other models may need to be developed as a benchmark for ongoing forecast evaluation.

Finally, when an appropriate model has been selected, further decisions may need to be made regarding the forecasting of exogenous economic variables and the protocol for integrating (and iterating) the macroeconomic and fiscal models.[48]

---

[48] Pike and Savage (1998) provide an overview of the way in which HMT macroeconomic model forecasts of taxes and tax bases are combined, compared, and constrained to the more detailed tax revenue forecasts of HMRC. Although the paper is dated, practitioners report that the process has remained largely unchanged (with the exception of scrutiny by the OBR).

Table 3: Model comparison

	Criteria	Rule	Univariate	Multivariate	VAR	ECM	Macro	DSGE	Microsim
Application	forecasting	good	good	fair	good	good	fair	poor	poor
	policy	fair	poor	good	poor	fair	fair	fair	good
Accuracy	short run	fair	good	fair	good	fair	fair	fair	N/A
	medium run	fair	fair	fair	fair	good	fair	fair	N/A
Communication	story telling	fair	poor	good	poor	good	good	fair	good
	transparency	good	fair	good	fair	fair	fair	poor	good
Data compatibility		fair	good	fair	fair	fair	fair	good	poor
Resources		good	good	fair	good	fair	poor	poor	poor

Legend

Rule: Forecasting by technical assumption (rule of thumb, growth accounting model, and external consensus)

Univariate: Univariate time series approaches

Multivariate: Multivariate regression models

VAR: Vector autoregressive models

ECM: Error-correction models

Macro: Large-scale macroeconometric models

DSGE: Dynamic stochastic general equilibrium models

Microsim: Microsimulation models

8. References

Addison-Smyth, D., McQuinn, K., & O'Reilly, G. (2008). Estimating the Structural Demand for Irish Housing (No. 1/RT/08). Central Bank of Ireland.

Alquist, R. L. Kilian, & R. J. Vigfusson. (2013). Forecasting the Price of Oil. *Handbook of Economic Forecasting*, 2A, edited by G. Elliott and A. Timmermann. Amsterdam, North-Holland: 427–507.

An de Meulen, P., Schmidt, & T., Micheli, M. (2011). *Forecasting House Prices in Germany*. Ruhr Economic Papers 294.

An, S., & Schorfheide, F. (2007). Bayesian analysis of DSGE models. Econometric reviews, 26(2-4), 113-172.

Auterson, T. (2014). Forecasting house prices. *Office for Budget Responsibility.* Working paper No. 6.

Barari, M., Sarkar, N., Kundu, S., & Chowdhury, K. B. (2014). Forecasting house prices in the United States with multiple structural breaks. International Econometric Review (IER), 6(1), 1-23.

Batchelor, 2007. Forecaster Behaviour and Bias in Macroeconomic Forecasts, Ifo Working Paper Series No. 39, Ifo Institute - Leibniz Institute for Economic Research at the University of Munich.

Belsky, E., Drew, R., & McCue, D. (2007). Projecting the underlying demand for new housing units: Inferences from the past, assumptions about the future. Joint Center for Housing Studies Harvard University. W07-7.

Bernanke, B. S., Boivin, J., & Eliasz, P. (2004). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach (No. w10220). National Bureau of Economic Research.

Best, M. C., & Kleven, H. J. (2013). Housing market responses to transaction taxes: Evidence from notches and stimulus in the UK. London School of Economics.

Brooks, C., & Tsolacos, S. (2010). Real estate modelling and forecasting. Cambridge University Press.

Brooks, C., & Tsolacos, S. (1999). The impact of economic and financial factors on UK property performance. Journal of Property Research, 16(2), 139-152.

Brown, J. P., Song, H., & McGillivray, A. (1997). Forecasting UK house prices: A time varying coefficient approach. Economic Modelling, 14(4), 529-548.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3), 307-327.

Box, G. E., & Jenkins, G. M. (1976). Time series analysis: forecasting and control, revised ed. Holden-Day.

Burgess, S., Burrows, O., Godin, A., Kinsella, S., & Millard, S. (2016). A dynamic model of financial balances for the United Kingdom. Staff Working Paper No. 614.

Caldara, D., R. Harrison, & Lipinska, A. (2012). Practical tools for policy analysis in DSGE models with missing channels. Finance and Economics Discussion Series 2012-72, Board of Governors of the Federal Reserve System (US).

Carnot, N., Koen, V., & Tissot, B. (2014). Economic forecasting and policy. Palgrave Macmillan.

Case, K. E., Quigley, J. M., & Shiller, R. J. (2005). Comparing wealth effects: the stock market versus the housing market. Advances in macroeconomics, 5(1).

Case, K. E., & Shiller, R. J. (1990). Forecasting prices and excess returns in the housing market. Real Estate Economics, 18(3), 253-273.

Case, K. E., & Shiller, R. J. (1988). The efficiency of the market for single-family homes. NBER Working Paper No. 2506.

Christiano, L. J., Eichenbaum, M., & Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. Journal of political Economy, 113(1), 1-45.

Crawford, G. W., & Fratantoni, M. C. (2003). Assessing the forecasting performance of regime-switching, ARIMA and GARCH models of house prices. Real Estate Economics, 31(2), 223.

Crespo Cuaresma, J. (2010), "Can Emerging Asset Price Bubbles be Detected?", OECD Economics Department Working Paper, No.772.

Del Negro, M., & Schorfheide, F. (2012). DSGE-based forecasting. Federal Reserve Bank of New York Staff Reports.

De Vroey, M. (2016). A history of macroeconomics from Keynes to Lucas and beyond. Cambridge University Press.

Dicks, M. J. (1990). A simple model of the housing market (No. 49). Bank of England, Economics Division.

Dixon, P. B., & Rimmer, M. (2009). Forecasting with a CGE model: does it work. Monash University, Centre of Policy Studies and the Impact Project.

Edwards, S. (2004). GENESIS: SAS based computing environment for dynamic microsimulation models. Mimeo, Department of Work and Pensions, London.

Elming, W., & Ermler, A. (2016). Housing equity, saving and debt dynamics over the Great Recession (No. W16/12). Institute for Fiscal Studies.

Enders, W. (2014). Applied economic time series. 4th edition. Iowa State University.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the Econometric Society, 987-1007.

Fullerton, T. M., Laaksonen, M. M., & West, C. T. (2001). Regional multi-family housing start forecast accuracy. International Journal of Forecasting, 17(2), 171-180.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). Bayesian data analysis. Texts in statistical science series.

Gervais, O., & Gosselin, M. A. (2014). Analyzing and Forecasting the Canadian Economy through the LENS Model. Bank of Canada.

Girouard, N., & André, C. (2006). Measuring cyclically-adjusted budget balances for OECD countries. Available at SSRN 2005002.

Granziera, E., Luu, C., & St-Amant, P. (2013). The Accuracy of short-run forecast Combinations. Canadian Economic Analysis Bank of Canada Review.

Granger C W J, & Newbold P. (1986) Forecasting Economic Time Series, 2nd edn. Academic Press, Orlando, FL.

Gupta, R., Jurgilas, M., & Kabundi, A. (2010). The effect of monetary policy on real house price growth in South Africa: A factor-augmented vector autoregression (FAVAR) approach. Economic modelling, 27(1), 315-323.

Hendry, D. F. (1984). Econometric modelling of house prices in the United Kingdom. Econometrics and quantitative economics, 211-252.

Herbst, E. P., & Schorfheide, F. (2015). Bayesian Estimation of DSGE Models. Princeton University Press.

Hicks, J. R. (1939). The foundations of welfare economics. The Economic Journal, 49(196), 696-712.

HM Treasury (2016). HM Treasury analysis: the long-term economic impact of EU membership and the alternatives. Londen, VK: HM Treasury, link.

Hyndman, R., & Athanasopoulos, G. (2012). Forecasting: principles and practice.

Iacoviello, M., & Neri, S. (2010). Housing market spillovers: evidence from an estimated DSGE model. American Economic Journal: Macroeconomics, 2(2), 125-164.

Kapetanios, G., Labhard, V., & Price, S. (2008). Forecast combination and the Bank of England's suite of statistical forecasting models. Economic Modelling, 25(4), 772-792.

Klein, L. R., & Goldberger, A. S. (1955). econometric model of the United States, 1929-1952.

Kydland, F. E., & Prescott, E. C. (1982). Time to build and aggregate fluctuations. Econometrica: Journal of the Econometric Society, 1345-1370.

Larson, W. D. (2011). Evaluating alternative methods of forecasting house prices: A post-crisis reassessment. Available at SSRN 1709647.

Li, J., & O'Donoghue, C. (2013). A survey of dynamic microsimulation models: uses, model structure and methodology. International Journal of Microsimulation, 6(2), 3-55.

Lis, A. (2013). Forecasting Canadian Housing Prices: Assessing the Forecasting Performance of ARIMA and GARCH Models.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions—five years of experience. Journal of Business & Economic Statistics, 4(1), 25-38.

Maitland-Smith, J. K., & Brooks, C. (1999). Threshold autoregressive and Markov switching models: an application to commercial real estate. Journal of Property Research, 16(1), 1-19.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2008). Forecasting methods and applications. John Wiley & Sons.

Mankiw, N. G., & Weil, D. N. (1989). The baby boom, the baby bust, and the housing market. Regional science and urban economics, 19(2), 235-258.

Manuelli, R. E., & Sargent, T. J. (1988). Longer Trading Periods in the Townsend Turnpike Model. Graduate School of Business, Stanford University.

Matthews, P. (2016). Forestalling ahead of property tax changes. Office for Budget Responsibility Working Paper No. 10.

McCue, D. (2009). Addendum to W07-7: Using long-term demand projections to estimate short-run market imbalances. Joint Center for Housing Studies Harvard University.

Meen, G., Gibb, K., Leishman, C. & Nygaard, C. A. (2016) Housing economics: a historical approach. Palgrave Macmillan.

Meen, G., (2013). Home-ownership for future generations. Urban Studies, 50 (4). pp. 637-656

Meen, G., (2012). Housing and the economy. The SAGE Handbook of Housing Studies, 251-273.

Meen, G. (2009). A Simple Model of Housing and the Credit Crunch. University of Reading, Department of Economics

Meen, G. (2001). Modelling Spatial Housing Markets: Theory, Analysis, and Policy.

Meen, G. (1990). The removal of mortgage market constraints and the implications for econometric modelling of UK house prices. Oxford Bulletin of Economics and Statistics.

Mian, A. R., Rao, K., & Sufi, A. (2013). Household balance sheets, consumption, and the economic slump. Chicago Booth Research Paper, (13-42).

Morand, E., Toulemon, L., Pennec, S., Baggio, & R., Billari, F. (2010). Demographic modelling: the state of the art. Paris, INED. FP7-244557 Projet SustainCity, 39.

Moro, A., & Nuño, G. (2012). Does total-factor productivity drive housing prices? A growth-accounting exercise for four countries. Economics Letters, 115(2), 221-224.

Newton, H.J. (1988). Timeslab: A Time Series Analysis Laboratory. Pacific Grove, CA: Wadsworth & Brooks/Cole Publishing.

O'Donoghue, C., Lennon, J., & Hynes, S. (2009). The Life-cycle Income Analysis Model (LIAM): a study of a flexible dynamic microsimulation modelling computing framework. International Journal of Microsimulation, 2(1), 16-31.

Office for Budget Responsibility. (2016). Devolved taxes forecast, November 2016.

Office for Budget Responsibility. (2013). The macroeconomic model. Briefing Paper No. 5.

Office for Budget Responsibility. (2012a). Forecasting Scottish taxes, March 2012.

Office for Budget Responsibility. (2012b). A small model of the UK economy. Working paper No. 4.

Office of the Parliamentary Budget Officer. (2016). Budget 2016: Key Issues for Parliamentarians. Canada.

Riddel, M. (2004). Housing-market disequilibrium: an examination of housing-market price and stock dynamics 1967–1998. Journal of Housing Economics, 13(2), 120-135.

Riddel, M. (2000). Housing market dynamics under stochastic growth: An application to the housing market in Boulder, Colorado. Journal of regional science, 40(4), 771-788.

Robidoux, B., & Wong, B. S. (1998). The Canadian Economic and Fiscal Model-1996 Version: Part 1-Model Structure (No. 1998-05). Department of Finance Canada.

Romer, P. (2016). The Trouble with Macroeconomics.

Rousová, L., & Van den Noord, P. (2011). Predicting peaks and troughs in real house prices. OECD Economics Department. ECO/WKP(2011)51.

Park, J., & Hong, T. (2011). Trends and prospects of the US housing market using the Markov switching model. Journal of Urban Planning and Development, 138(1), 10-17.

Pike, T. & Savage, D. (1998). Forecasting the Public Finances in the Treasury. Fiscal Studies vol. 19, no. 1, pp. 49–62.

Poterba, J. M. (1984). Tax subsidies to owner-occupied housing: an asset-market approach. The quarterly journal of economics, 729-752.

Sargent, R. G. (1987). An overview of verification and validation of simulation models. In Proceedings of the 19th conference on Winter simulation (pp. 33-39). ACM.

Sargent, T. J., & Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. New methods in business cycle research, 1, 145-168.

Sauerbier, T. (2002). UMDBS-a new tool for dynamic microsimulation. Journal of Artificial Societies and Social Simulation, 5(2).

Scottish Fiscal Commission. (2015). Report on Draft Budget 2016-17. December.

Slanicay, M. (2014). Some Notes on Historical, Theoretical, and Empirical Background of DSGE Models. Review of Economic Perspectives, 14(2), 145-164.

Slobodyan, S., & Wouters, R. (2012). Learning in an estimated medium-scale DSGE model. Journal of Economic Dynamics and control, 36(1), 26-46.

Slutzky, E. (1937). The summation of random causes as the source of cyclic processes. Econometrica: Journal of the Econometric Society, 105-146.

Smets, F., & Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. The American Economic Review, 97(3), 586-606.

Smets, F., & Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the euro area. Journal of the European economic association, 1(5), 1123-1175.

Stock, J. (2002). Time series: economic forecasting. International Encyclopedia of the Social Sciences.

Stock, J.H. & M.W. Watson (2002), Forecasting Using Principal Components from a Large Number of Predictors, Journal of the American Statistical Association 97, 1167 – 1179.

Stevenson, S., & Young, J. (2014). A multiple error-correction model of housing supply. Housing Studies, 29(3), 362-379.

Stevenson, S., & Young, J. (2007). Forecasting housing supply: Empirical evidence from the Irish Market. European Journal of Housing Policy, 7(1), 1-17.

Tong, H. (1983). Threshold Models in Non-Linear Time Series Analysis. New York, Springer-Verlag.

Tsolacos, S. (2006). An assessment of property performance forecasts: consensus versus econometric. Journal of Property Investment & Finance, 24(5), 386-399.

Wold, H., (1938). A Study in the Analysis of Stationary Time Series, Stockholm: Almgrist & Wiksell.

Wolfson, M., & Rowe, G. (1998). LifePaths–Toward an integrated microanalytic framework for socio-economic statistics. In 26th General Conference of the International Association for Research in Income and Wealth, Cambridge, UK.

Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between Time-Series? A study in sampling and the nature of time-series. Journal of the royal statistical society, 89(1), 1-63.

Van den Noord, P. (2006). Are House Prices Nearing a Peak? A Probit Analysis for 17 OECD Countries. Organisation for Economic Co-operation and Development.

9. Acronyms

ARIMA   Auto-regressive integrated moving average
BVAR   Bayesian vector autoregression
COMPASS   Central Organising Model for Projection Analysis and Scenario Simulation.
DFM   Dynamic factor modelling
DSGE   Dynamic stochastic general equilibrium
ECM   Error-correction mechanism
FRBNY   Federal Reserve Bank of New York
GARCH   Generalised autoregressive conditional heteroskedasticity
HMT   HM Treasury
LBTT   Land and Buildings Transaction Tax
LENS   Large Empirical and Semi-structural model
MultiMod   The IMF's multicountry macro model
OBR   Office for Budget Responsibility
OLS   Ordinary Least Squares
RICS   Royal Institution of Chartered Surveyors
ToTEM   Terms-of-Trade Economic Model
VAR   Vector autoregression

Appendix – model assessment summaries

Table A1: Technical assumptions

Criteria		Result
Application	forecasting	Good. There are examples of all three technical assumptions used to forecast housing prices and transaction volumes: rules of thumb, growth accounting models, or consensus forecasts.
	policy	Fair. Univariate models forecast how prices and transactions typically move over time, but do not consider the underlying causes. Neither of the three technical assumptions excel at fiscal costings. There is some scope for adjusting assumptions for fiscal sensitivity tables or scenario analysis. Alternative capacity for fiscal impact costing would need to be developed.
Accuracy	short run (quarters one to eight)	Fair. Evidence suggests consensus forecast averaging limits can produce better forecasts than any one model on its own. Rules of thumb and growth accounting are useful for unpredictable series; however, there is sufficient evidence that more sophisticated models are useful for forecasting housing markets.
	medium run (years three to five)	Fair. There is little difference between rules of thumb and growth accounting models for the short and medium run. Some evidence suggests that the consensus average forecast is less accurate than simple univariate approaches and regression models.
Communication	story telling	Fair. Technical assumptions vary in their ability to tell a story. Generally, they are not tied to economic theory in a way that permits forecast revisions and the path of the forecast to be well articulated. However, assumptions such as the growth rate of GDP or the consensus survey can capture general economic trends.
	transparency	Good. All three approaches can be made independent and transparent (provided the consensus forecast is viewed as an externally-imposed rule. However, the underlying methodologies and assumptions used to produce the individual forecasts underlying the consensus would typically not be available.
Data compatibility		Fair. Rule of thumb and growth accounting frameworks work well with quarterly data and are generally well suited for medium-run forecasts. However, few Scotland-specific external forecasts of the housing market are produced on a quarterly basis for a five-year time horizon.
Resources		Good, with a qualification: internal capacity may still be required for policy analysis. However, for forecasting purposes, a technical assumption is well suited for a small tax as a share of the overall budgetary revenues. Easily estimated or imposed in spreadsheets and statistics software packages.
Evidence		Alquist and Vigfusson (2013), An de Meulen et al. (2011), Auterson (2014), Batchelor (2007), Belsky et al. (2007), Case and Shiller (1987), Crawford & Fratantoni (2003), Granziera et al. (2013), Makridakis (2008), McCue (2009), An de Meulen et al. (2011), Moro and Nuño (2012), Tsolacos (2006).

Table A2: Univariate and time series models

Criteria		Result
Application	forecasting	Good. Univariate models can be used for both prices and transactions. They can produce *ex ante* forecasts without needing to be conditioned on auxiliary forecasts of exogenous variables.
	policy	Poor. Because they use their own history—with no links to other economic or fiscal variables—univariate models are not suitable for policy analysis by themselves. They may, however, be used as a component of a model that can perform policy analysis, such as forecasting the parameters of the housing price distribution to impose on the housing base.
Accuracy	Short run (quarters one to eight)	Good. Evidence suggests that in the short-run, univariate time series models routinely beat other forecasting procedures and models with more structure.
	Medium run (years three to five)	Fair. Evidence and theory suggests that as the forecast horizon lengthens, univariate models are likely to be relatively less reliable than models grounded in theory and equilibrium concepts. The length of the useful forecast horizon is determined by the speed of decay, which is rarely greater than 8 quarters; however, their ability to capture simple trends beyond two years may prove sufficient for the housing market.
Communication	story telling	Poor. The models are mostly a black box without explanatory variables or theory to provide interpretable coefficients.
	transparency	Fair. Models are estimated by applying relatively little judgment. Model equations can be published; however, they are usually updated frequently and coefficients and specification may change faster than documentation. Because the model structure is data driven and not intuitive, it could pose a challenge for budget scrutiny.
Data compatibility		Good. Univariate models work well with the number of observations of quarterly data available for the Scottish housing market.
Resources		Good. Univariate models require few resources and little expertise to estimate. Similar to current methodology and may not require more resources. Easily estimated in modern software packages such as Eviews and Stata. Cumbersome in Excel.
Evidence		Barari et al. (2014), Bollerslev (1986), Box and Jenkins (1976), Brooks and Tsolacos (1999), Brown et al. (1997), Crawford and Fratantoni (2003), Engle (1982), Enders (2014), Fullerton et al. (2001), Hyndman (2012), Lis (2013), Maitland-Smith and Brooks (1999), An de Meulen et al. (2011), Park and Hong (2012), Stock (2002), Stevenson and Young (2007), Tong (1983).

Table A3: Multivariate regression models

Criteria		Result
Application	forecasting	Fair. Potential to produce forecasts which closely match the true tax base. Many examples in the literature for prices and transactions. However, because they rely on exogenous variables, they cannot produce *ex ante* forecasts, reducing their score.
	policy	Good. Can include a wide range of variables representing government policy and other explanatory variables that are well-suited for policy analysis, if they are specified with that in mind.
Accuracy	short run (quarters one to eight)	Fair. Because they emphasize theory, they can miss important dynamics and useful information in the data that a theoretical specification for causal inference may miss. Can include lags of explanatory variables to improve short-run dynamics.
	medium run (years three to five)	Fair. Similar to short-run. Accuracy of medium run depends on accuracy of exogenous variables.
Communication	story telling	Good. Multivariate models are well-positioned to provide a narrative for the forecast path and forecast revisions. Explanatory variables are intuitive, parameter signs fit with economic theory, and causal interpretations are often possible.
	transparency	Good. Multivariate models can be estimated with little judgment beyond which variables to include and the specification. Model equations can be published. Relationships are grounded in theory and unlikely to change. If they change it would be in magnitude of coefficients, not sign or overall specification. Model specification is intuitive and permits useful debate among budget stakeholders and fiscal watchdogs.
Data compatibility		Fair. Multivarite models work well with the number of observations of quarterly data available for the Scottish housing market. There may be some limitations on potential explanatory variables.
Resources		Fair. Few analysts with some specialised knowledge. A small increase in complexity compared to the current model process. May require more than current resources. Easily estimated in Excel and modern software packages such as Eviews and Stata.
Evidence		Auterson (2014), Bank of England (2008), Carnot et al. (2014), Case and Shiller (1987), Dicks (1990), Fullerton et al. (2001), Makriditis et al. (2008), Meen (2013, 2012, 2009, 2001, 1990), Poterba (1984), Tsolacos (2006).

Table A4: Vector autoregressive models

Criteria		Result
Application	forecasting	Good. Examples have been found in the literature for both prices and transactions. They can produce *ex ante* forecasts without needing to be conditioned on auxiliary forecasts of exogenous variables.
	policy	Poor. VARs are constructed under the principle of parsimony—the lowest number of variables that are significant. If a variable is not useful, then it should not be included. VARs generally lack causal economic and fiscal links to do policy analysis. Some shocks (impulse response) analysis possible.
Accuracy	short run (quarters one to eight)	Good. Under ideal circumstances, VARs would offer superior forecasting potential, particularly for short-run forecasts. This is conditional on having sufficient Scottish historical data that doesn't limit the VAR's specification.
	medium run (years three to five)	Fair. Evidence and theory suggests that as the forecast horizon lengthens, VAR models are likely to be relatively less reliable than models grounded in theory and equilibrium concepts. The length of the useful forecast horizon is determined by the speed of decay.
Communication	story telling	Poor. The models are mostly a black box without a causal specification grounded in theory. Often have counterintuitive signs and the long lag structure makes interpretation difficult.
	transparency	Fair. Models are estimated mechanically by applying relatively little judgment. Model equations can be published; however, they are usually updated frequently and coefficients and specification may change faster than documentation. Coefficients tend not to be intuitive and could pose a challenge for budget scrutiny.
Data compatibility		Fair. Possible to estimate, depending on number of explanatory variables and lags included, but small sample size makes difficult to estimate a model. Is suitable for quarterly forecasting if seasonal dummies are included or deseasonalised data is used. Can produce 5-year forecast, but performance is better for short-run.
Resources		Good. VAR models require few resources and modest expertise to estimate (with modern software). Similar to current methodology and may not require more resources. Easily estimated in software packages such as Eviews and Stata. Cumbersome in Excel.
Evidence		Bernanke et al, (2004), Brooks and Tsolacos (2010), Sargent and Sims (1977), Stock and Watson (2002).

Table A5: Error-correction models

Criteria		Result
Application	forecasting	Good. Examples have been found in the literature for both prices and transactions. They can produce limited *ex ante* forecasts depending on specification.
	policy	Fair. Not ideal for policy analysis, but depending on specification and integration in wider macroeconomic model can do impulse responses and limited policy analysis.
Accuracy	short run (quarters one to eight)	Fair. ECMs are generally regarded as offering good short-run forecasting potential.
	medium run (years three to five)	Good. Because they are based on equilibrium concepts, ECMs are generally regarded as more capable at five-year horizons than other approaches.
Communication	story telling	Good. ECMs are generally interpreted easily, signs and magnitudes are meaningful, and a consistent narrative can be formed.
	transparency	Fair. Equations can be published and their specification (particularly model coefficients) is unlikely to change frequently. Model equations are more opaque to budget scrutinizers with only a general economics background than more simple regression equations, but the equations are nonetheless more economically intuitive than VARs.
Data compatibility		Fair. Possible to estimate, but low sample size for parameter confidence, especially given the crisis. The current approach is likely more appropriate for the sample size. Can handle quarterly frequency data.
Resources		Fair. Can be estimated and maintained with relatively few analysts. But specialized knowledge and advanced degree in economics or statistics recommended.
Evidence		Addison-Smyth et al. (2008), Auterson (2014), Enders (2014), Riddel (2000, 2004), Stevenson and Young (2014)

Table A6: Large-scale macroeconometric models

Criteria		Result
Application	forecasting	Fair. Generally, not well-specified for best forecasts or correspondence to the true tax base. But can be complemented by a richer suite model framework to arrive at a rich specification for both prices and transactions.
Application	policy	Fair. Not suited for fiscal impact costing. Can assess fiscal multipliers, responses of government intervention, dynamic scoring. Often on a different accounting basis (national accounts rather than commercial accounting standards).
Accuracy	short run (quarters one to eight)	Fair. Near-term is based on monitoring; however, because national accounts data is subject to revisions, the starting point and recent history may not be pinned down. Rely on dynamics more than the multivariate econometric models of Subsection 3.3, but generally still grounded in theory.
Accuracy	medium run (years three to five)	Fair. Because of its theoretical underpinnings and extensive reliance on long-run equilibrium conditions (closing of the output gap) and frequently levels variables (with heavy reliance on error correction models), these models are likely to improve upon naïve forecasts for the medium run. However, medium run forecast will suffer from the same concerns as the short run.
Communication	story telling	Good. Can produce consistent, intuitive narratives in-line with economic theory of the housing market and its impact on the macroeconomy. Coefficient signs make sense.
Communication	transparency	Fair. Estimated with considerable judgment, especially for smoothing the near term (first two quarters) and squaring with economic monitoring. Model equations can be published. Relationships are grounded in theory and unlikely to change. If they change it would be in magnitude of coefficients, not sign or overall specification. Model specification is intuitive and permits useful debate among budget stakeholders and fiscal watchdogs.
Data compatibility		Fair. The Quarterly National Accounts of Scotland should provide enough information for a simple model of GDP; however, the specification of the housing sector (residential investment) and accompanying aggregate behavioural equations is likely to be rudimentary.
Resources		poor. One or two very experienced and skilled analysts is sufficient to maintain a large-scale macroeconometric model once developed; however, the development is a significant undertaking. Useful to have several less experienced analysts in charge of economic monitoring and to work independently on model components such as import/export modules. Would require programmable statistics software beyond Excel.
Evidence		Auterson (2014), Gervais and Gosselin (2014), Carnot et al. (2014), Case et al. (2005), Robidoux and Wong (1998), De Vroey (2016), Granger and Newbold (1986), Hicks (1939), HM Treasury (2016), Elming and Erlmler (2016), Klein and Goldberger (1955), Mankiw and Weil (1989), Manuelli and Sargent (1988), Meen (2012), OBR (2013), PBO (2016), and Romer (2016).

Table A7: Dynamic stochastic general equilibrium models

Criteria		Result
Application	forecasting	Poor. While some versions are suited to forecasting, few examples have been found in the literature for housing prices and transactions. Very unlikely could be specified well enough to forecast the LBTT tax base. Can perform *ex ante* forecasts: conditional variables jointly determined within the system.
	policy	Fair. DSGE models may have use for modelling the transmission through the economy of housing market scenarios. DSGE models are not appropriate for static policy costing (costings that are estimated using a single market and do not consider the feedback effects of the rest of the economy). However, they can be used for dynamic scoring (modelling the impulse of government revenue changes through the wider economy).
Accuracy	short run (quarters one to eight)	Fair. Evidence is mixed and depends on focus (small models generally outperform larger models) and specification.
	medium run (years three to five)	Fair. Some evidence to suggest that DSGE models may offer better performance in the medium run than other models less grounded in economic theory
Communication	story telling	Fair. Theoretical foundations lend themselves to intuitive narratives. However, complexity of interactions runs significant risks of becoming a 'black box' with difficult or no interpretation.
	transparency	Poor. Results and assumptions can be opaque and harder to explain to policymakers. Considerable judgment is applied throughout estimation. External scrutiny would require specialist training.
Data compatibility		Good. A DSGE model could be estimated using the Quarterly National Accounts of Scotland. Small DSGE models can be constructed on as few as seven data series; however, incorporating the household sector would require the addition of two or three others. All are available.
Resources		Poor. Although be more compact and take less effort than a large-scale macroeconomic model. Most institutions with DSGE models maintain a full suite of other models to complement and compare with the DSGE's results, as it is generally too aggregated to provide the detail required for budgeting and policy. It still requires forecasting exogenous variables, monitoring for the current period and near short-run.
Evidence		An and Schorfheide (2007), Burgess et al. (2016), Caldara, Harrison and Lipinska (2012), Carnot et al. (2011), Christiano et al. (2005), De Vroey (2016), Del Negro and Schorfheide (2012), Herbst and Schorfheide (2015), Iacoviello and Neri (2010), Kydland and Prescott (1982), Romer (2016), Sargent (1987), Slanicay (2014), Slobodyan and Wouters (2012), Smets and Wouters (2003), Smets and Wouters (2007).

Table A8: Microsimulation models

Criteria		Result
Application	forecasting	Poor. Microsimulation models cannot in themselves produce a forecast for the housing market. They require a model to forecast the baseline assumptions (growth factors) for variables that determine the evolution of receipts. However, they are useful for forecasting the distribution of the prices over time, given these inputs.
	policy	Good. Microsimulation models are ideal for assessing the impact of different tax rates and thresholds on government revenues. A drawback is that they are by nature mechanical and require ad hoc adjustments to the simulation output to incorporate behavioural responses.
Accuracy	short run (quarters one to eight)	N/A. As the model itself cannot forecast, this criterion is not applicable. However, if economic variables forecast with auxiliary models are accurate, then the microsimulation model should give accurate conditional estimates of revenues.
	medium run (years three to five)	N/A. ibid
Communica tion	story telling	Good. Because microsimulation models are a rote imposition of the tax code on real households, communication of microsimulation results are intuitive.
	transparency	Good. The underlying equations are mechanical identities, and aside from weights to scale results to the population level, little estimation and no judgment is applied.
Data compatibility		Poor. There are limitations for forecasters' access to tax-payer level data in Scotland. Microsimulation model development may not be possible currently, but could be considered in the future if alternative data protocols are explored.
Resources		Poor. Microsimulation models are likely to require considerable resources to develop and maintain.
Evidence		Best and Klevin (2013), Edwards, (2004), O'Donoghue (2013), OBR (2016), Morand et al. (2010), Sauerbier (2002), Wolfson and Rowe (1998).

social research

CSR
GOVERNMENT SOCIAL RESEARCH
Social Science in Government